

Contracting for Health: Evidence from Cambodia*

Indu Bhushan,¹ Erik Bloom,² David Clingingsmith,³
Rathavuth Hong,⁴ Elizabeth King,² Michael Kremer,³
Benjamin Loevinsohn,² and J. Brad Schwartz⁵

August 24, 2007

Abstract

Distortions in both government healthcare provision and private fee-for-service provision are severe in developing countries, leading many to question whether the large increases in health care spending that have been proposed to have much impact. Cambodia recently tried a new approach to service delivery by contracting out management of government health services to NGOs in five districts that had been randomly made eligible for the approach. The contracts specified targets for maternal and child health service improvement. Targeted outcomes improved dramatically by about 0.5 standard deviations relative to comparison districts. Changes in non-targeted outcomes were small. The

*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 thank Dan Choate and Russell Weinstein for excellent research assistance. 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.

¹Asian Development Bank

²World Bank

³Harvard University

⁴ORC Macro

⁵University of North Carolina—Chapel Hill

program reduced provider absence, greatly improved targeted outcomes, 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.

1 Introduction

Many have called for increased public expenditures on health in the developing world (World Health Organization 2001). Others question whether increased expenditure or foreign aid to the health sector can be effective in such circumstances, citing the lack of a cross-country correlation between public health spending and child mortality (Filmer & Pritchett 1999). Even within countries such as India, having easy access to government health facilities seems to have no effect on infant mortality (Chaudhury *et al.* 2005). Health care systems in developing countries seem to artfully combine the worst aspects of government and private provision. Provider 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. Weak public health care systems lead even poor patients to turn to private providers. Private practitioners' incentives are strong, but they are often not well aligned 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). Many private providers lack any medical training. Banerjee *et al.* (2004) found that private medical practitioners in rural Rajasthan gave injections to 68% of patients, IV drips to 12%, but tests only to 3%. An estimated 30%-

50% of prescriptions written in India are unnecessary or contraindicated (Das & Sánchez-Páramo 2002; Phadke 1998). Moreover, private health care provision leaves consumers exposed to substantial financial risk from adverse health shocks. Xu *et al.* (2003) found that 5% of Cambodian households suffer a financially catastrophic health shock each year, and that across countries the occurrence of such shocks are positively correlated with the fraction of medical spending that is out-of-pocket.

Starting in 1999, Cambodia tried an alternative approach to health service provision. The government tendered management of government health services in certain districts for contract to private bidders. 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. They were responsible for services at district hospitals, sub-district health centers, and more remote health posts. Public health expenditure increased to pay for the bids. Performance was explicitly measured against eight service delivery indicators. Inadequate performance could lead to sanctions and would reduce the likelihood that the contract would be renewed.

From a theoretical perspective, district-level contracting 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. On the other hand, 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.

Empirical estimation of program impact is eased by the random selection of districts for tendering of contracts, but is hampered by two difficulties. First, while eight districts were randomly selected from a set of 12 for tendering of bids, technically acceptable bids were received and the program implemented in only five of these districts. We therefore estimate the causal effects of the program by using the initial random assignment of tender 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 the twelve-unit randomization. We compute standard errors using both the Huber-White clustered estimator and 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 targeted service outcomes of about one-half of a standard deviation. To cite two examples, the program increased receipt of vitamin A by children under 5 by 21 percentage points and receipt of antenatal care by pregnant women by 33 percentage points. The project improved the management of government health centers, particularly in the availability of 24-hour service, the presence of staff, supervisory visits, and the presence of supplies and equipment. The program

did not have large effects on health service indicators not explicitly mentioned in the contract. There is some limited evidence the program improved self-reported health. People in program districts reduced visits to untrained service providers such as drug sellers and traditional healers and increased curative care visits to public facilities. Reduced out-of-pocket spending on curative care offset increased public spending, so the program did not increase, and probably decreased, overall health spending. The program likely reduced individual risk from health shocks.

We draw on several previous studies of the contracting project (Bhushan *et al.* 2002; Keller & Schwartz 2001; Schwartz & Bhushan 2004). The previous work focused on contracted outcomes; we improve on the measurement of these outcomes and contextualize them within a multitasking framework. The previous studies used a 1997 baseline survey and 2001 midterm survey and found that targeted outcomes improved more in districts than in comparison districts. However, the 2001 midterm survey did not collect data in the three districts initially assigned to treatment for which acceptable bids were not received. Previous work therefore could not take advantage of the initial randomization of treatment eligibility to estimate causal impacts. Its 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, program effects could be overestimated. To compute causal estimates, we conducted an additional survey in 2003 that covered all of the districts initially randomized into treatment. This chapter also differs from earlier work in accounting for the

cluster-randomized nature of the design, in looking at families of outcomes using average effect size (AES) techniques, and in presenting new evidence on health center management, non-targeted outcomes, expenditures, and perception of the quality of care.

The remainder of the chapter 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 empirical methods. Section 5 estimates project impact on targeted outcomes, non-targeted outcomes, careseeking behavior, and final health outcomes. Section 6 discusses the impact of the program on health center management and perception of care quality. Section 7 discusses the effects of the program on public, private, and overall health care expenditures. Section 8 concludes the chapter.

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. Section 2.1 discusses health care in the Cambodian context. We describe the program and randomization in sections 2.2 and 2.3. Section 2.4 describes variants of the program and how it was financed. Section 2.5 examines some of the human resource management practices undertaken by contractors under the program. Section 2.6 describes the data collected about

program outcomes.

2.1 Background on Health Care in Cambodia

The Khmer Rouge regime killed nearly 20% of Cambodia's population during the late 1970s. When the Vietnamese invaded in 1979, only 50 doctors were left in the country. Under the Vietnamese-backed regime that ruled the country from 1979 until 1993, private medical practice was banned, and public facilities were sparse and poorly equipped. Little rural health infrastructure existed, and investment was low. In 1993, the Vietnamese left. Cambodia held elections and moved toward a market economy. NGOs established a large presence in the country, including the health sector, though their geographical footprint was limited. Private medical practice, now legal, boomed. Nearly all trained medical providers in the country worked for the government, but many public employees ran private practices on the side. Absenteeism in public facilities was high, and diversion of patients to private practice was common. Public sector salaries for health staff averaged only about 85% of Cambodia's US\$283 per-capita GDP in 1998 (Conway 2000). Sale of pharmaceuticals was de facto unregulated, with most vendors having no pharmaceutical or medical training.

While the country made tremendous improvements to public health between 1993 and 1997, the contracting projects 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 less than two years of age were fully immunized and only 9% of pregnant women had antenatal care. About one third of

sick individuals who sought curative care went only to a drug seller.

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 in the public sector depends in part on political work and political connections.

Traditional understandings of health and disease are strong in Cambodia and affect individual care choices and the practices of medical providers. Thirty-nine percent of mothers whose under five-year-old child died reported supernatural forces as primary cause of death in a 2000 study (RACHA 2000). A medical anthropology field study suggests that for ordinary rural Cambodians, the perceived quality of medical care is closely tied to being dispensed drugs (Van de Put 1991). Diagnostic skill and medical expertise are not viewed as critical, 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. Patients are reported take a favorable view of practitioners who listen to instructions on which drugs to administer. Practitioners commonly give medically unnecessary vitamin injections and glucose drips that may make patients feel better in the short run.

Fighting finally ended in Cambodia in 1998, and shortly afterward the contracting project was introduced. The project was part of program that also included a large national expansion of rural health centers.

2.2 The Contracting of Health Services Pilot Project

The contracting project ran from 1999 to 2003 and covered about 1.26 million people, or about 11% of Cambodia's population (Cambodia 2002; Schwartz & Bhushan 2004). The contracting approach was then expanded to additional districts, though not as a randomized experiment.

Contractors were required to provide the standard bundle of care services mandated for all health districts in Cambodia, known as the minimum package of activities. The contracts set explicit targets for 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. The program goals were between 160% and 450% of the baseline levels and were to be achieved within four years (Table 1). 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. 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.

2.3 Selection of Districts and Contract Award System

The twelve districts selected to participate in the project came from three provinces in south-central Cambodia—Takeo, Prey Veng, and Kampong Cham. Districts that

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

Indicator	Definition	Program Goal %	Comparison	Baseline (%)	
				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 contraception.	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 treatment status.

did not contain the provincial capital or receive of other significant development assistance at baseline were chosen to be eligible for treatment.

Each of 12 eligible districts was randomly assigned. Eight districts were assigned to the contracting and four to comparison. Two variants of contracting were tried in four districts each; the differences between variants are detailed in section 2d below. The 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 two districts to be eligible for contracting and one district to be in the comparison group. This left two districts to be assigned in Kampong Cham and one to be assigned in Prey Veng. They were randomly assigned later in Phnom Penh.

Management of government health care services in the districts randomized to eligibility was put out for competitive bid by qualified organizations 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. 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 defined 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 contractors management plan.

The contracts were awarded in late 1998. Only five of the eight districts randomized into the treatment group were contracted. In two districts, no technically responsive bids were received and in a third, they were too expensive. Districts that did not receive technically responsive bids tended to have higher baseline levels of the targeted outcomes than successfully contracted districts, while the district that received bids that were too expensive had lower baseline levels than successfully contracted districts.

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. Contractors took over management of the five successfully contracted districts by April 1999. The comparison districts and those districts not successfully contracted continued to be managed by the local employees of the Ministry of Health.

2.4 Treatment Variants, Project Finance, and Monitoring

The project was originally designed to test two variants of the contracting approach that differed in the degree of control to be given to the contractors and the process by which budgets were decided. These distinctions became blurred in practice. The variants were called contracting-in and contracting-out.

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 staff; they could request transfers but not hire or fire. The 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 district management had 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. They could implement the pay structure of their choosing. Existing Ministry of Health staff in the contracting-out districts could join the contractors organization and take leaves of absence from the civil service. If the contractor decided to fire them, they would be transferred to a government post in a different district. In practice, only a few staff members in contracting out districts were fired.

The contracting-out districts received their funds directly from the ADB. 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. Most of the comparison districts proposed to utilize the budget supplement to conduct immunization outreach activities. Audit irregularities were often encoun-

tered in three of the four comparison districts, sometimes resulting in the suspension of payments. The amounts actually used by comparison districts were often less than the full \$0.25 per capita. Districts were allowed to use the operating supplement for salaries after contracting-in staff refused to work under the government pay structure. The comparison districts were also given health care management consulting services and management training as part of the Basic Health Services Project, which the treatment districts did not receive. Those districts randomized into treatment that were not successfully contracted received neither the supplement, nor management consulting, nor management training (see Table 2).

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, and in one case payment was suspended for one quarter, until the problem was remedied.

2.5 Management Practices of the Contractors

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. 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

Table 2: Actual Treatment Characteristics of Project Districts

	Contracting Out	Contracting In	Comparison	Not Successfully Contracted
Number of Districts	2	3	4	3
Staffing	Hired at market rates. MOH staff could take leave of absence	MOH staff on government salary, usually given performance	MOH staff on government salary, often given supplement from user charges	MOH staff on government salary, often given supplement
Procurement	NGO responsible	NGO responsible but through MOH system	Through MOH system	Through MOH system
Budget Supplement	No	Yes	Yes	No
Technical Assistance & Management Training	No	No	Yes	No
Contracting Project Payments	\$3.98	\$1.05	\$0.00	\$0.00
Contracting Budget Supplement	\$0.00	\$0.31	\$0.05	\$0.00
Other Basic Health Services Project Spending	\$0.09	\$0.13	\$0.08	\$0.09
Government	\$0.20	\$1.20	\$1.37	\$1.16
NGO/Donation	\$0.06	\$0.19	\$0.06	\$0.10
Cost Recovery	\$0.10	\$0.14	\$0.03	\$0.01
TOTAL excluding Cost Recovery	\$4.42	\$3.02	\$1.59	\$1.36
Percentage of Comparison	278%	190%	100%	86%

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

regulations without salary supplements. The two contracting-out districts paid fixed salaries considerably higher than the government salary, 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. Contracted districts had better managerial supervision and better staff attendance than comparison districts (see section 6a).

Soeters & Griffiths (2003) describe the operation of the contracting-in model by HealthNet International (HNI) Pereang district. The contractor viewed staff motivation as a key challenge, and implemented a performance-based incentive system. It subcontracted with managers of health centers and the hospital, who in turn established contracts with their staff. 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 and monitoring out-of-pocket patient expenditure. Service prices were set at about 60% of the prevailing market price. User fees thus collected paid about half of staff incentives, with the other half coming through the contracting program. HNI conducted its own surveys and spot checks.

All of the successful bidders hired expatriates for some management and advisory roles, with an average of one expat per district at any given time. They typically filled the role of district manager, overseeing between about 60 and 120 local staff. Roughly half of the expatriate staff were Asians and half were Westerners. Annual salaries for expat managers ranged widely between about US\$15,000 and US\$60,000. Expats may have had several advantages over locally born staff that lead to contractors employing them. Expats are likely less subject to local political pressures than

locals. They take part in career ladders in international organizations in which their incentives to perform well are much steeper than those faced by locals. They also may bring different management views and capabilities than those available locally. It is possible that expat managers played an important role in the success of the program. The generalizability of the program to a different setting may depend on whether a similar pool of talent is available. Its important to note that international NGOs often have a large presence in post-conflict environments such as Cambodia during the contracting program.

2.6 Data

A baseline survey was conducted in 1997 before interventions were made, and a full follow-up was conducted in 2003. Data on individual health care outcomes and care-seeking behavior was collected from 30 randomly-chosen villages in each district. About 20,000 individuals in 3,700 households were surveyed in each survey round. The same villages were sampled in both survey years, but within villages a new random sample of households was taken each time. 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.

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. This idea can be formalized in a simplified Holmstrom-Milgrom 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 T and NT and the effort types e_1

and e_2 and let them be produced as follows

$$T = f(e_1, e_2) + \epsilon \quad (3.1)$$

$$NT = g(e_1, e_2) + \eta. \quad (3.2)$$

The agent cares about compensation w as well as the cost of exerting effort $c(e_1, e_2)$,

$$u(w, e_1, e_2) = w - c(e_1, e_2) \quad (3.3)$$

The agent is paid a linear wage in the amount of the contracted outcome produced

$$w = \alpha + BT. \quad (3.4)$$

The agent's first order conditions are

$$\frac{\partial c}{\partial e_1} = B \frac{\partial f}{\partial e_1} \quad (3.5)$$

$$\frac{\partial c}{\partial e_2} = B \frac{\partial f}{\partial e_2} \quad (3.6)$$

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 T output that results from effort. Increasing B will typically increase T . A change in effort that results in an increase in T may either increase or decrease NT depending on whether T and NT are complements or substitutes. For example, immunizing children requires health workers to make field visits. While in

the field, they may treat cases of diarrhea that would have otherwise gone untreated. These two activities are complements. On the other hand, health workers are absent from their clinics during field visits, which may negatively affect the services provided to walk-in patients.

This simple framework does not consider the possibility that the NGO and the 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

We use the initial randomization of treatment as an instrumental variable for actual treatment to produce 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 (section 4.1). Computing standard errors requires care given that randomization was at the district level. We use both the techniques of clustering at the district level and of randomization inference (section 4.2). We also compute the average effect of contracting across a family of related outcomes that comprise a domain of interest, such as outcomes explicitly targeted by the program for improvement (section 4.3).

4.1 Estimation of per-comparison causal effects

We report treatment-on-treated (TOT) causal effects (Angrist *et al.* 1996; Imbens & Angrist 1994). The TOT measures the effect on a health care outcome of being in a district that received the contracting treatment. The TOT is a local average treatment effect (LATE), meaning that it is the expected effect on an individual whose district would get the treatment were it randomized into the experiment.

The basic regression TOT model for the effect of the program on a particular outcome y_{idpt} for an individual i from district d and province p at time t is:

$$y_{idpt} = \beta_0 + \beta_1 I_d^{CON} + \beta_2 I_t^{2003} + \beta_3 (I_d^{CON} \times I_t^{2003}) + p_{pt} + \epsilon_{idpt}. \quad (4.1)$$

The dummy variable I_d^{CON} indicates whether district d actually received the contracting program. The TOT effect of contracting is given by β_3 . The province \times year fixed effects p_{pt} reflect the quasi-stratification of the randomization by province and absorb time-varying shocks at the province level. We use a dummy variable for treatment eligibility I_d^{ELIG} as an instrumental variable for actual treatment status.

Instrumental variables estimation requires that the instrument I_d^{ELIG} be uncorrelated with the error term ϵ_{idpt} . This means that the outcome of the randomization procedure can only have an effect on the outcome through the treatment. The contracting program does not satisfy this requirement in that the districts that were eligible for treatment but not contracted did not get the same budget supplement or technical assistance received by the comparison districts. This places a downward bias on IV estimates of the contracting TOT effect β_3 .

Two notes of caution are warranted in interpreting the results. First, constants in the regressions should not be interpreted as values for the comparison districts because of the inclusion of province \times year effects. Second, while the estimates are unbiased (consistent), they are subject to sampling error. Since the IV estimate is a multiple of the difference between the group randomly assigned to treatment and that assigned to comparison, if districts selected for treatment do very well or very badly, the resulting IV estimate may seem to imply that the level of the dependent variable in the treated districts has risen above one or fallen below zero even if this is impossible.

In some specifications we will want to distinguish between the effects of the two contracting program variants, contracting-in and contracting-out. To do so we use a version of equation 4.1 in which the contracting dummy I_d^{CON} is replaced with separate dummies for each variant I_d^{CON-IN} and $I_d^{CON-OUT}$. The coefficients on the interactions of these two dummies with the post-treatment dummy I_t^{2003} provide estimates of the treatment effects for each variant.

4.2 Randomization at district level

Our analysis examines the effects of district-level independent variables on individual-level dependent variables. Individual outcomes may be correlated due to a district-level time varying variables, such as might arise with 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 & Lang (2001); Wooldridge (2003) have

pointed out that asymptotic justification of this estimator assumes a large number of aggregate units. Simulations in Bertrand *et al.* (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.

We also report hypothesis test generated using randomization inference (Rosenbaum 2002). This approach involves generating placebo random assignments P_j and associated regression coefficients β_j . Denote by $\{P_j\}$ the set of all possible assignments from the randomization process described in section 2.3. There are 6,480 unique random assignments, which are equally likely to occur. Suppose we are interested in the effect of the actual assignment P_T on outcome Y_{ij} . Let this true effect be β_T . Now consider β_j in the following regression equation:

$$Y_{ij} = \delta + \beta_j P_j + \nu_{ij} \tag{4.2}$$

Since P_j is a randomly generated placebo, $E(\beta_j) = 0$. Let $F(\widehat{\beta}_j)$ be the empirical c.d.f. of $\widehat{\beta}_j$ for all elements of $\{P_j\}$. We can now perform a hypothesis test by checking if the measured treatment effect is in the tails of the distribution of placebo treatments. We can reject $H_0 : \beta_T = 0$ with a confidence level of $1 - \alpha$ if $\widehat{\beta}_T \leq F^{-1}(\frac{\alpha}{2})$ or $\widehat{\beta}_T \geq F^{-1}(1 - \frac{\alpha}{2})$. Since the placebo assignments only vary across clusters, this method takes intracluster correlations into account. We then compute placebo treatment effects for each of these random assignments using the placebo version of the ITT estimating equation 4.1. We compute a p-value by noting where the true effect lies in the distribution of placebo effects.¹

¹We can use the ITT p-values to assess the null hypothesis of no TOT effect because under this

We present hypothesis tests based both on clustering and on randomization inference because randomization inference 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 observations, two in the treatment group and four in the comparison group. With minimal restrictions on the error term, such as single-peakedness, one would reject the null hypothesis of no effect if the difference between the treatment and comparison group averages were large enough relative to the differences within the groups. However, since there are 15 combinations in this set of placebo randomizations $\{P_j\}$, each with $1/15$ probability of being selected, randomization inference will 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. It is always impossible to reject the hypothesis that errors take the form $-x$ with probability $\frac{1}{3}$ and $\frac{x}{2}$ with probability $\frac{2}{3}$, where x is the difference between groups.

4.3 Summary measures of causal effects

We have very low power in testing individual effects. Looking at families of related outcomes together can help increase power to the extent that outcomes within families are not perfectly correlated.

Following Kling *et al.* (2004); Logan & Tamhane (2003); O'Brien (1984), suppose that there are K outcomes in a family and each is scaled so that the treatment effects π_k are positive if they are desirable. Consider the measure of average effect size τ

 null the effect of taking up treatment is the same as the effect of not taking it up (Greevy *et al.* 2004).

over the family of K outcomes $\{y_{dtk}|k = 1 \dots K\}$ in which each treatment effect is normalized by the standard deviation σ_k of the change in the outcome Δy_{dtk} ,² where

$$\tau = \frac{1}{K} \sum_{k=1}^K \frac{\pi_k}{\sigma_k}. \quad (4.3)$$

O'Brien (1984) showed that, under the assumption of a constant treatment effect across outcomes within the family, could be used to test the one-sided hypothesis $H_0 : \lambda = 0$ against $H_A : \lambda > 0$, where $\frac{\pi_k}{\sigma_k} = \lambda \forall k$.

Because the 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 \bar{y}_{vdtk} . The joint equation is estimated by instrumental variables, again using random assignment as an instrument for treatment

$$\begin{pmatrix} \bar{y}_{vdt1} \\ \vdots \\ \bar{y}_{vdtK} \end{pmatrix} = (I_K \otimes W)\theta + \varphi_{vdtk} \quad (4.4)$$

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

²Alternative 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, its possible 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 outcomes.

5 Health Services and Final Health Outcomes

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. Section 5.1 considers the eight health-service indicators explicitly targeted by the program. Sections 5.2, 5.3, and 5.4 consider non-targeted outcomes, curative care seeking behavior, and final health outcomes.

Baseline data for the 21 outcomes we examine in the chapter suggest treatment and comparison groups were comparable at baseline (see Tables 3, 6, 7, 8, and 12). The baseline level of one outcome is statistically significant at 5% under clustering and randomization inference; one additional outcome is significant at 10% under clustering. Under random assignment, we would expect one significant coefficient at 5%. The baseline levels are therefore consistent with random assignment of treatment.

5.1 Targeted outcomes

Before plunging into statistical analysis of program impact, this subsection first presents histograms on targeted outcomes. We then discuss then per comparison and average effect estimates of the effect of treatment. To show robustness, we also report an alternative estimation that controls for household wealth.

Figures 1 and 2 show histograms of the percentage point changes in district level average outcomes over the course of the program. The tremendous overall improvement in Cambodia over the period is immediately apparent. The strong secular increase in service provision that forms the background for the treatment

effects was driven 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.

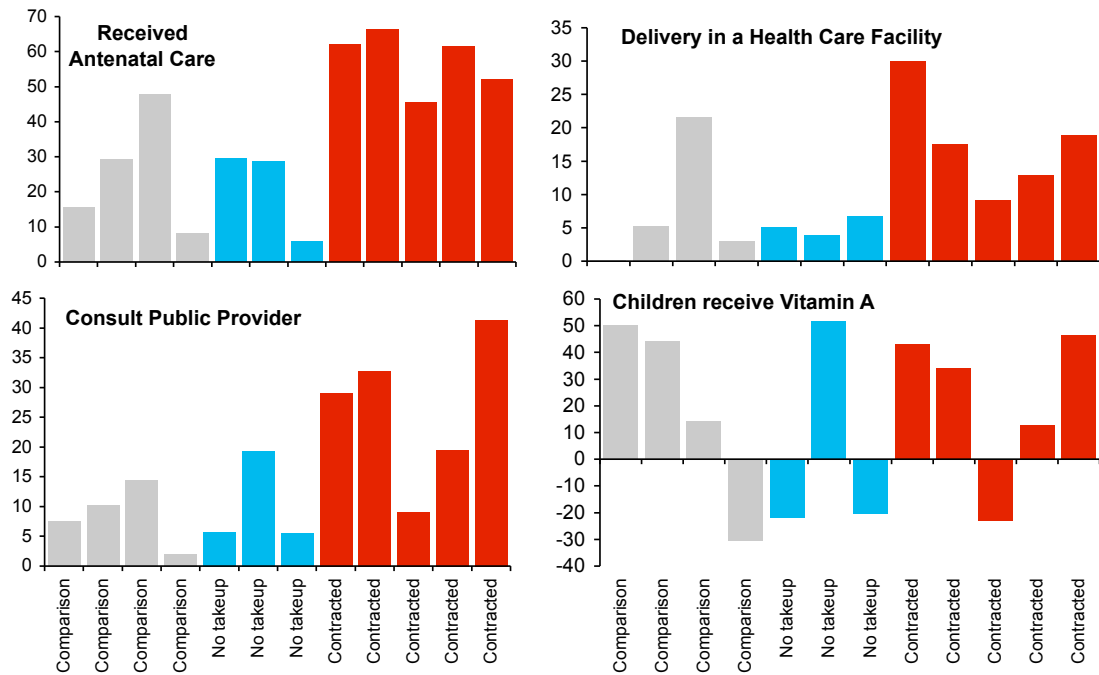


Figure 1: Changes in Contracted Outcomes by District 1997-2003, Part 1

It is also clear from the histograms 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.

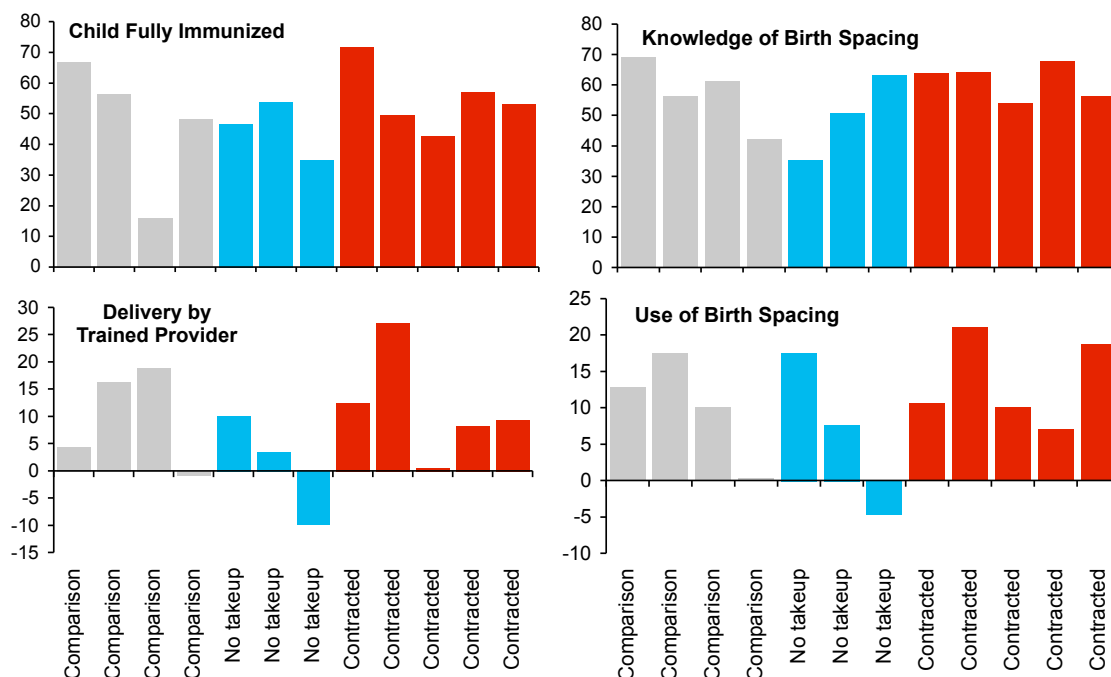


Figure 2: Changes in Contracted Outcomes by District 1997-2003, Part 2

Changes in comparison districts and not-taken-up districts are roughly similar. Note that the large differences between actually treated and not-treated districts will show up as large standard errors in the regression. The outcomes are mostly binary, so that in cases where improvement of the outcome was very strong for both treatment and comparison, such as knowledge of contraception, we will have very little power to detect an effect. TOT effects are large for most of the eight targeted outcomes (Table 3). Contracting treatment led to a statistically significant 22 percentage point increase in the choice of a public sector facility when needing a curative care consultation, very substantially above the 4% baseline level for the comparison group

and much larger than the 8 percentage point secular increase between 1997 and 2003. The program caused a 33 percentage point increase (statistically significant at 10% under randomization inference) in the receipt of antenatal care by pregnant women, compared to a baseline level of 9% in the comparison group and over and above a 26 percentage point secular increase. Vitamin A receipt by children was increased by 21 percentage points (statistically significant at 10% under randomization inference) by the program against a baseline comparison ratio of 43%.

Contracting improved targeted outcomes an average of about 0.52 comparison-group standard deviations (Table 3, final column). This is a very large average effect. The hypothesis of no average positive effect is rejected at 1% under clustering and 5% under randomization inference. The average effects were not different across program variants. 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. The TOT estimates tend to be slightly larger than estimates that follow the methodology used in previous analyses of the project, which omitted those eligible districts where treatment was not taken up (Table 4). An OLS estimate of the average effect of contracting on targeted outcomes is 0.50 standard deviations, compared to the causal estimate of 0.52 standard deviations. The average effect was 0.60 standard deviations for the lower half of the wealth distribution and 0.47 standard deviations for the upper half of the wealth distribution, though we fail to reject the null hypothesis of no difference.

It is worth considering results from an alternative specification that seeks to

Table 3: TOT Effects on Changes in Targeted Outcomes

	Full Immunization	Vitamin A Antenatal Care	Trained Delivery	Delivery in Facility	Use Contraception	Know Contraception	Use Public Facilities	Average Effect Size
Contracted District	-0.100 (0.09)	-0.060 (0.04)	0.059 (0.11)	0.019 (0.02)	0.042 (0.05)	0.004 (0.07)	-0.005 (0.02)	
Contracted District X 2003	0.143 (0.09)	0.205** (0.09)	-0.004 (0.05)	0.103 (0.07)	0.037 (0.05)	0.011 (0.07)	0.220*** (0.03)	0.516*** (0.07)
Randomization inference p-value	0.32	0.08	0.98	0.12	0.57	0.89	0.01	0.04
Year 2003	0.298**	0.184**	0.332**	0.117	0.137***	0.598***	0.153***	
Constant	(0.10)	(0.06)	(0.11)	(0.07)	(0.04)	(0.05)	(0.02)	
	0.509***	0.465***	0.136***	0.054***	0.095*	0.135**	0.024*	
	(0.10)	(0.03)	(0.11)	(0.01)	(0.04)	(0.05)	(0.01)	
Observations	5,100	11,213	4,993	4,976	6,994	9,537	11,223	
R-squared	0.275	0.128	0.248	0.03	0.019	0.338	0.12	
Contracting In X 2003 effect	0.139	0.091***	0.364***	0.057	0.118***	0.077**	0.176***	0.505***
Contracting Out X 2003 effect	0.15	0.417***	0.263***	-0.123*	0.074	-0.038	0.289***	0.537***
Comparison mean 2003	0.81	0.61	0.34	0.10	0.23	0.80	0.13	
Comparison mean 1997	0.34	0.43	0.24	0.03	0.13	0.22	0.04	
H ₀ : CO=CI, p-value								0.83

Notes: IV regressions including province \times year fixed effects. Average effects are differential increases caused by treatment in units of standard deviations of changes in the outcomes. Standard errors in parentheses corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. Average effects tests null of zero effect against two-sided alternative.

Table 4: OLS Estimates Comparing Contracted Districts to Comparison Districts

	Full Immunization	Vitamin A Care	Antenatal Delivery	Trained Delivery	Delivery in Facility	Use Contraception	Know Contraception	Use Public Facilities	Average Effect Size
Contracted District	-0.113 (0.07)	-0.036 (0.03)	-0.016 (0.03)	0.027 (0.09)	0.004 (0.01)	-0.002 (0.01)	-0.054 (0.05)	-0.016** (0.01)	
Contracted District X 2003	0.157** (0.06)	0.165* (0.07)	0.317*** (0.07)	0.021 (0.03)	0.105 (0.06)	0.043 (0.03)	0.053 (0.04)	0.190*** (0.03)	0.499*** (0.06)
Clustering SE	0.288*** (0.08)	0.210*** (0.06)	0.340*** (0.09)	0.166*** (0.03)	0.116 (0.07)	0.134*** (0.03)	0.570*** (0.04)	0.172*** (0.03)	
Constant	0.517*** (0.09)	0.449*** (0.03)	0.151*** (0.02)	0.319** (0.10)	0.064*** (0.01)	0.123*** (0.01)	0.173*** (0.03)	0.030*** (0.01)	
Observations	3,826	8,357	3,751	3,751	3,737	5,254	7,157	8,433	
R-squared	0.30	0.13	0.30	0.05	0.08	0.03	0.37	0.14	
Comparison mean 2003	0.81	0.61	0.35	0.34	0.10	0.23	0.80	0.13	
Comparison mean 1997	0.34	0.43	0.09	0.24	0.03	0.13	0.22	0.04	

Notes: Regressions include only treated and comparison districts; eligible districts that were not treated are not included. Regressions including province \times year fixed effects. Average effects are differential increases caused by treatment in units of standard deviations of changes in the outcomes. Standard errors in parentheses corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. Average effects tests null of zero effect against two-sided alternative.

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 5 shows specifications that include controls for household assets in the per-comparison results. This will not be the 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 the 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 the 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.

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 5, Panel B). Controlling for assets improves the fit of the per-comparison TOT regressions slightly (Table 5, Panel A). Overall, the size of the per-comparison effects is similar to regressions that do not control for assets. The average TOT effects of contracting on targeted outcomes controlling for assets is a statistically significant 0.59 comparison group standard deviations. It is reassuring that effects are similar under this specification.

Table 5: Robustness Check: Wealth Controls

Panel A: TOT Estimates										
	Full Immunization	Vitamin A Antenatal Care	Trained Delivery	Del. in Facility	Use Contraception	Know Contraception	Use Public Facilities	Average Effect Size		
Contracted District	-0.097 (0.08)	0.008 (0.04)	0.063 (0.09)	0.020 (0.02)	0.046 (0.05)	0.008 (0.07)	-0.003 (0.02)			
Contracted District X 2003	0.146 (0.08)	0.204** (0.09)	0.334*** (0.05)	0.109 (0.06)	0.046 (0.05)	0.012 (0.07)	0.218*** (0.03)	0.594*** (0.09)		
Clustering inference p-value	0.33	0.07	0.91	0.07	0.50	0.92	0.00	0.04		
Year 2003	0.263** (0.10)	0.178** (0.07)	0.273** (0.10)	0.100** (0.04)	0.117** (0.04)	0.577*** (0.05)	0.152*** (0.03)			
Constant	0.462*** (0.09)	0.452*** (0.03)	0.078** (0.03)	0.210** (0.10)	0.074 (0.04)	0.115* (0.05)	0.017 (0.01)			
Observations	5,084	11,178	4,979	4,962	6,975	9,510	11,191			
R-squared	0.28	0.13	0.27	0.09	0.08	0.34	0.12			
Contracting In X 2003 effect	0.141*	0.091	0.368***	0.067	0.124*	0.085	0.173***	0.579***		
Contracting Out X 2003 effect	0.157	0.412***	0.267	-0.110	0.078	-0.028	0.075	0.288***		
Comparison mean 2003	0.81	0.61	0.35	0.34	0.10	0.23	0.80	0.13		
Comparison 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

Panel B: 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
Clustering SE	(0.02)	(0.06)	
Randomization inference p-value	0.23	0.46	

Notes: IV regressions include province \times year fixed effects and wealth controls. Standard errors 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. Average effects tests null of zero effect against two-sided alternative.

5.2 Non-targeted outcomes

In section 3 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.³ 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 positive but statistically insignificant effects on non-targeted outcomes (Table 6). Contracting had a statistically significant positive TOT effect of 21 percentage points on knowledge of HIV risk factors. Overall, contracting had a positive average effect on non-targeted outcomes of about 0.22 comparison-group standard deviations, though the effect was not significant. The average effect of contracting on non-targeted outcomes was less than half as large as the effect on

³The 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 is important for its immune system. Water should not be given to babies under six months old because of the danger of disease transmission.

targeted outcomes. These results alleviate the concern that the incomplete nature of explicit contracts would lead contracting to have large negative effects on non-targeted outcomes.

5.3 Care-seeking behavior

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 examine care-seeking behavior for all household members during the past month (see Table 7). 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 these, about 48% saw drug sellers and traditional healers (grouped as unqualified private providers), 44% saw qualified providers in private practice, and 8% saw qualified providers in a public setting. Contracting reduced the chance that someone visited any provider by 4.1 points, though the effect was not statistically significant. The contracting treatment made individuals about 4.3

Table 6: TOT Effects on Changes in Nontargeted 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
Contracted District	-0.050 (0.05)	0.433 (0.43)	-0.102** (0.04)	0.029 (0.03)	0.005* (0.00)	-0.038 (0.04)	
Contracted District X 2003	0.061 (0.05)	0.936 (0.57)	0.111 (0.07)	-0.012 (0.09)	0.056 (0.04)	0.208** (0.08)	0.215 (0.17)
Clustered SE							
Randomization inference p-value							
Year 2003	0.31 (0.071)	0.38 (2.805***)	0.34 (0.181***)	0.96 (0.429***)	0.20 (-0.084***)	0.02 (0.264***)	0.25
Constant	(0.06)	(0.60)	(0.05)	(0.09)	(0.03)	(0.06)	
	0.867*** (0.04)	0.754* (0.39)	0.658*** (0.03)	0.050 (0.03)	0.995*** (0.00)	0.250*** (0.03)	
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	
Contracting In X 2003 effect	0.018	1.119**	0.180*	0.015	-0.037	0.211**	0.281
Contracting Out X 2003 effect	0.144*	0.578	-0.029	-0.064	-0.093	0.196	0.087
Comparison mean 2003	0.93	2.79	0.77	0.35	0.95	0.42	
Comparison 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 \times 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.

percentage points more likely to consult a qualified public provider compared to a comparison group baseline of 0.6%. This effect is statistically significant at 1%. Contracting reduced visits to unqualified providers by about 6.1 percentage points (significant at 10%) against a baseline of 8.5%.

We compute the average effect of contracting on careseeking behavior in two ways. In both cases, we treat visits to biomedically trained providers as positive and visits to drug sellers and traditional healers as negative. 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.

In the first computation, we take an individual not consulting a provider to be positive. Those who report being ill almost always consult a provider, so most of those who did not consult a provider were not ill. Contracting improved careseeking outcomes by an average of 0.34 standard deviations. The effect was significant at 10% under randomization inference. We condition the second analysis on whether someone visited a provider. Contracting increased visits to a trained and qualified provider by a statistically significant 0.51 comparison-group standard deviations. We fail to reject the null hypothesis that average effects for the contracting-in and contracting-out variants 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, the following subsection will show that patients had a poorer

Table 7: TOT Effects on Changes in Care-Seeking Behavior

	Was any provider consulted in the past month?			Average Effect Size	
	None	Unqualified Provider	Qualified Private Provider	Qualified Public Provider	Unconditional on Visiting a Provider
Contracted District	-0.047 (0.05)	0.039 (0.03)	0.004 (0.02)	0.005 (0.01)	
Contracted District X 2003	0.041 (0.04)	-0.061* (0.03)	-0.020 (0.02)	0.043*** (0.01)	0.335* (0.17)
Clustered SE	0.39	0.06	0.60	0.00	0.505*** (0.13)
Randomization inference p-value					0.00
Year 2003	-0.064* (0.03)	0.045* (0.02)	-0.01 (0.02)	0.055*** (0.01)	
Constant	0.846*** (0.04)	0.047* (0.02)	0.115*** (0.02)	0.005 (0.00)	
Observations	54,062	54,062	54,062	54,062	
R-squared	0.01	0.01	0.01	0.02	
Contracting In X 2003 effect	0.003	-0.042*	0.015	0.036***	0.251*
Contracting Out X 2003 effect	0.118	-0.098	-0.090*	0.054***	0.481*
Comparison mean 2003	0.826	0.081	0.080	0.031	0.551**
Comparison mean 1997	0.824	0.085	0.095	0.009	
H ₀ : CO=CI, p-value					0.31
					0.88

Notes: IV regressions with province \times year 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 conditions on seeking care from any provider.

perception of the quality of care in program districts.⁴

5.4 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 the 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 be 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 had no statistically significant effects on the incidence of illness generally, on the incidence of diarrhea in children, or on the survival of a child less than one year of age (see Table 8). Overall, contracting had a positive though statistically insignificant average effect of 0.18 standard deviations on final health. The contracting out variant had a positive average effect 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.

⁴It 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.

Table 8: TOT Effects on Final Health Outcomes

	No illness during past month	No diarrhea past month (child under 5)	Child <1 Alive	Average Effect
Contracted District	-0.047 (0.05)	-0.048 (0.07)	0.027 (0.02)	
Contracted District X 2003	0.047	0.079	-0.020	0.180
Clustered SE	(0.04)	(0.07)	(0.02)	(0.23)
Randomization inference p-value	0.41	0.44	0.46	0.73
Year 2003	-0.061* (0.03)	0.05 (0.05)	0.011 (0.02)	
Constant	0.824*** (0.05)	0.723*** (0.06)	0.957*** (0.01)	
Observations	54,062	9,850	4,930	
R-squared	0	0.02	0.01	
Contracting In X 2003 effect	-0.001	-0.01	-0.011	-0.045
Contracting Out X 2003 effect	0.145	0.252	-0.043	0.616**
Comparison mean 2003	0.185	0.26	0.97	
Comparison mean 1997	0.202	0.35	0.97	
H ₀ : CO=CI, p-value				0.26

Notes: IV regressions with province \times year 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.

6 Management under Contracting

The contracting treatment led to large increases in targeted outcomes. We would like to understand more about the management practices contractors engaged in as a result of the program. These practices may shed more light on how the contractors improved service outcomes.

6.1 Health center management

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. (No baseline survey was conducted.)

Contracting dramatically increased availability of services at health centers (see Table 9). The two treatment variants had statistically different average effects, so we show estimates separately for contracting-in and contracting-out. 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.) Contracting increased the probability that round the clock services at the health center; the 83-point contracting-in effect was statistically significant. Both variants had large point estimates for whether an unannounced visit to a health center would find the center open and treating patients, and that all scheduled staff would be present.⁵ The effects on staff presence were statistically significant under clustering for both variants, while the contracting-in effect on the health center being open was significant under clustering.

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

⁵Unfortunately we do not have more detailed information on the fraction of staff present.

Table 9: TOT Effects for Health Facility Management

	Facilities, Staffing, and Equipment										Services	
	Permanent health center building open	24 hour service at health center	Unann. visit: open w/ patients	Unann. visit: All sched staff present	Health center equipment index	Health center supplies index	All child vaccs available at health center	Number outreach last month	Delivery services scheduled	Delivery services offered?		
Contracting In--Treated	0.236**	0.826***	0.477**	0.496**	3.530***	5.531***	-0.155*	-2.690	0.193**	0.246		
Clustered S.E.	(0.08)	(0.11)	(0.22)	(0.17)	(0.66)	(1.37)	(0.08)	(2.06)	(0.07)	(0.16)		
Randomization inference, p-value	0.22	0.03	0.17	0.13	0.10	0.10	0.48	0.32	0.02	0.11		
Contracting Out--Treated	0.170	0.467	0.711	0.787***	2.990*	8.863**	0.146	3.414	0.139	0.403		
Clustered S.E.	(0.22)	(0.27)	(0.44)	(0.24)	(1.37)	(3.10)	(0.18)	(3.19)	(0.12)	(0.36)		
Randomization inference, p-value	0.60	0.46	0.25	0.11	0.29	0.13	0.84	0.61	0.45	0.15		
Observations	143	121	143	143	143	143	143	143	124	143		
R-squared	0.23	0.57	0.52	0.43	0.33	0.38	0.3	0.16	0.02	0.02		
Comparison Mean	0.74	0.21	0.45	0.24	15.02	25.02	0.36	14.31	-0.06	0.52		
	Supervision and Financing											
	Num. supervisor visits in 3m	Last visit: discuss MOH progs	Last visit: discussed problems	Registers match HIS reports	User fee system with posted fees	User fee income (2003 US\$)	Health center support from other NGOs?	Average Effect Size				
Contracting In--Treated	0.028	0.102	0.090	0.308	0.238**	93.925	-0.061	0.599***				
Clustered S.E.	(0.49)	(0.09)	(0.08)	(0.19)	(0.09)	(82.83)	(0.31)	(0.14)				
Randomization inference, p-value	0.96	0.25	0.23	0.17	0.12	0.35	0.87	0.04				
Contracting Out--Treated	5.654***	0.197	-0.123	0.127	0.284*	92.345	0.245	1.128***				
Clustered S.E.	(1.34)	(0.19)	(0.18)	(0.36)	(0.15)	(81.63)	(0.80)	(0.38)				
Randomization inference, p-value	0.08	0.21	0.52	0.72	0.38	0.28	0.78	0.04				
Observations	143	112	116	143	108	89	143	HO: CO=CI, p-value				
R-squared	0.51	0.12	0.13	0.25	0.17	0.19	0.2					
Comparison Mean	2.52	0.77	0.81	0.67	0.77	92.83	0.45					

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 in the average effect 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 fee system with posted fees, 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.

index were positive for both program variants and were statistically significant at 10% for contracting-in. Contracting-in showed an increase in the equipment index of 3.5 against a comparison mean of 15.0 and an increase of 5.5 in the supplies index against a comparison mean of 25.0.

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 supervisors 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.

Contracting increased the likelihood a health center would implement a user-fee system with posted fees. The contracting-in effect of 23.8 points was significant at 5% under clustering and the contracting-out effect of 28.4 points was significant at 10% under clustering. Posting user fees makes it more difficult for staff members to overcharge patients. Both variants had positive point estimates on the amount of user fee income, though they were not statistically significant.

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 9).⁶ Both contracting-in and contracting-out had statistically

⁶The 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

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 ($p < 0.01$); we conclude the while both variants substantially improved management, contracting-out did so even more than contracting-in.

6.2 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 opinion about the quality of the services at the health center/outreach? Respondents were asked to give their opinion about: 1) staff attitude and behavior (honest, polite, and caring), 2) technical competence of staff, 3) supply of medicine and equipment (quantity and quality), and 4) total cost of services.

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 per-

center was open and seeing patients and whether all scheduled staff were present, availability of child delivery service, whether a health center has a user fee system with fees 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.

centage 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 point estimates were only negative for the richer half. Point estimates were also negative for outreach (not shown), though none were statistically significant.

Given the results 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. 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. Discussions with contracting managers revealed they believed they had a different view of appropriate care than typical Cambodian providers, who may

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
Clustered 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**
Clustered 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
Constant	0.828***	0.710***	0.680***	0.879***	
	(0.05)	(0.05)	(0.04)	(0.05)	
Observations	2526	2499	2479	2524	
R-squared	0.07	0.13	0.12	0.01	
Contracted District	-0.127**	-0.103	-0.078*	0.021	-0.170*
Comparison Mean	0.63	0.50	0.50	0.87	
H ₀ : CO=CI					0.38

Notes: 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.

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. Finally, residents may have had higher expectations for services managed by an NGO than those managed by the regular government system.

7 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. Section 7.1 shows the project probably led to increases in total public spending on health. We will then show in Section 7.2 that the program decreased out-of-pocket spending by individuals on curative care procedures. In Section 7.3, we put 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. Lastly, we will examine non-experimental estimates of the effects of spending on targeted outcomes in Section 7.4 that suggest changes in public health spending have little independent effect on

service delivery.

7.1 Public health spending

We examine administrative data on public funds at the district level. Table 11 presents a breakdown of district-level public spending by major funding sources for comparison, contracting in, and contracting out districts. The data on public expenditures cover 1999 through 2003, and are based on the actual receipt of funds at the district level. 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 the latter half of 1999. 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 activities that used budget supplements in comparison districts were subject to auditing irregularities. As a consequence, comparison districts often did not receive the entire \$0.25 per capita per year for which they were eligible.

Total government per capita health spending for all of Cambodia was US\$2.88 during the course of the program.⁷ This figure includes spending at the district

⁷Based on (Cambodia 2004) and (World Bank 2007).

Table 11: Average Public Per Capita Spending by Source, 2000-2003 (USD 2003)

	Comparison	Contracting In	Contracting Out
District-level spending			
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 district-level spending excluding Cost Recovery	1.59	2.56	2.94
Percentage of Comparison		161%	185%
National-level spending			
National hospitals	0.60	0.60	0.60
National institutions and programs	0.85	0.85	0.85
Central ministry	0.43	0.43	0.43
TOTAL district- and national-level spending excluding Cost Recovery	3.47	4.44	4.82
Percentage of Comparison		128%	139%

Notes: We exclude public revenue derived from user fees from the public spending data since it is captured in the out-of-pocket spending data. National level spending is an estimate for 2001.

and national levels. Cambodia spent approximately US\$1.89 per capita, or 66% of its health budget, on activities undertaken at the national level. In 2001, these activities broke down as follows: national hospitals: 32%; central ministry: 23%; and national institutions and programs: 45%. Table 11 combines district- and national-level spending estimates into an overall estimate of per-capita public spending for the project districts.

Some of these national-level activities may not be relevant to the contracted districts, either because the contracted districts did not have access to them or because they duplicated activities performed by the contractors. For example, some of the administrative functions of the central government may have been duplicated in the contracting organizations. We explore two contrasting assumptions: 1) all districts benefit equally from the national-level spending, and ought therefore to be charged equally at US\$1.89 per capita for these expenses and 2) contracted districts derive no benefits from the national-level spending and ought not to be charged for it.

Using the first assumption, a TOT regression shows that contracting out increased public health spending by a statistically significant and very substantial \$3.64 per capita in 2003, against a comparison mean of \$1.59 (Table 12, column 4). Contracting in had a smaller effect of \$1.76 that was not significant under randomization inference. Under the second assumption, the TOT effect of the program on the level of public spending is much smaller but still positive for contracting out districts, though statistically insignificant (Table 12, column 5). The point estimate for contracting in is negative and statistically insignificant. 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.

7.2 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. 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.

Table 12: TOT Effects on Private and Public Health Care Spending

	Annualized Private Individual Curative Care Spending			Average Annual Public Health Spending Per Capita 2000-03	
	All Individuals excluding 0.5% tails (1)	All Individuals excluding 0.5% those who spent >US\$100 last month (2)	All Individuals excluding (3)	National-level spending equally distributed (4)	National-level spending allocated to non-contracted districts only (5)
CI-Treated	2.324 (06.71)	0.102 (03.06)	-0.082 (02.24)		
CO-Treated	47.286** (16.83)	17.003* (8.03)	10.507 (5.93)		
CI-Treated X 2003	-3.321	-0.521	-0.711	1.760**	-0.130
Clustered S.E.	0.50	0.81	0.71	(0.56)	(0.56)
Randomization inference p-value	0.95	0.98	0.96	0.27	0.84
CO-Treated X 2003	-55.856***	-22.122**	-15.608*	3.639***	1.749*
Clustered S.E.	0.00	0.01	0.03	(0.88)	(0.88)
Randomization inference p-value	0.01	0.01	0.02	0.10	0.18
Year 2003	3.186 (07.76)	-2.259 (03.67)	-2.458 (02.89)		
Observations	54,062	53,529	53,948	12	12
R-squared	0.01	0.01	0.01	0.89	0.62
Comparison mean 2003	\$12.12	\$7.51	\$9.84	\$1.87	\$3.73
Comparison mean 1997	\$18.76	\$10.42	\$15.17		

Notes: IV regressions with province \times year 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.

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 12, columns 1 to 3). The contracting out TOT effect is a large and statistically significant negative \$55.86 (2003 USD) using 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 \$47.26.⁸ 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.⁹ 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

⁸Throughout this chapter 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.

⁹We 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.

quantiles of the private expenditure distribution (Table 13). 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 government clinics, its plausible the strong negative effect for this variant may have been due to lower patient payments in public facilities.

7.3 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 Table 14. In panel A, national-level spending is treated as equally distributed, while in panel B it is not assumed to apply to districts under contracting. Column 1 of panel A 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. The parallel regression in column 1 of panel B, in which national-level spending is assumed to not apply

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

	Mean	90th percentile	95th percentile	99th percentile
CI-Treated	0.484 (3.54)	0.000 (7.70)	0.000 (18.95)	42.656 (82.10)
CO-Treated	21.832*** (3.56)	18.546** (7.71)	37.092** (15.10)	310.464*** (97.95)
CI-Treated X 2003	-1.012	-0.604	-2.416	-57.758
Clustered p-value	0.84	0.92	0.85	0.41
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 p-value	0.00	0.00	0.00	0.00
Randomization inference p-value	0.01	0.02	0.06	0.01
Year 2003	-7.324 (5.00)	-1.59 (5.08)	-26.463** (12.22)	-266.483*** (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: IV regressions with province \times year 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.

to contracted districts, shows a much smaller positive coefficient on contracting in. Columns 2 and 3 trim the distribution of private spending as discussed above. In panel A, the coefficients on contracting in and contracting out naturally are closer to zero, and still small and insignificant. In panel B, contracting has negative point estimates on the trimmed data.

The program shifted total spending into the public sector, which reduced individual financial risk from health shocks. Health spending in Cambodia is highly variable. In the 1997 baseline study, 75% of individuals had no out of pocket health in the previous month. The 95th percentile of this distribution was US\$6.80 and the 99th percentile is US\$46.37. Cambodia's monthly per capita GDP was US\$24.20.

7.4 Health spending and outcomes

While the impact of the program on public health spending was large, the macro literature on the effect of government spending on outcomes suggests spending alone might not have a large effect on outcomes. Using national level cross-sectional data Filmer & 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 1/7th of 1 percent.

We look at the OLS effect of district-level average public spending on both targeted and non-targeted outcomes, controlling for the presence of the program (Table 15). While we need to be cautious in attaching a causal interpretation to these results, they suggest that an additional dollar of public health spending has a small

Table 14: TOT Effects on Total Annualized Health Care Spending

Panel A: National-level public spending equally distributed

	2003 (Public Avg. 2000-03) (1)	2003 (Public Private excl.0.5% tails) (2)	2003 (Public Private <\$100 last month) (3)
CI-Treated	2.008	1.004	1.406
Clustered SE	(4.08)	(2.51)	(2.66)
Randomization inference p-value	0.56	0.65	0.57
CO-Treated	-5.654	-1.370	-1.361
Clustered SE	(6.49)	(3.99)	(4.24)
Randomization inference p-value	0.40	0.74	0.79
Constant	23.238*** (3.98)	17.994*** (2.45)	15.532*** (2.60)
Observations	12	12	12
R-squared	0.65	0.74	0.69
Comparison mean 2003	\$15.93	\$13.51	\$11.36

Panel B: National-level public spending allocated to non-contracted districts only during program

	(1)	(2)	(3)
CI-Treated	0.118	-0.886	-0.484
Clustered SE	(4.08)	(2.51)	(2.66)
Randomization inference p-value	0.90	0.72	0.88
CO-Treated	-7.544	-3.260	-3.251
Clustered SE	(6.49)	(3.99)	(4.24)
Randomization inference p-value	0.26	0.36	0.44
Constant	23.238*** (3.98)	17.994*** (2.45)	15.532*** (2.60)
Observations	12	12	12
R-squared	0.62	0.70	0.63
Comparison mean 2003	\$15.93	\$13.51	\$11.36

Notes: IV regressions with province \times year 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 on outcomes. We fail to reject the null hypothesis that additional spending has no effect on either targeted or non-targeted health service delivery outcomes.

8 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.

It is difficult to assess to what extent health services could be improved simply by spending more money in the existing public sector. We think simple spending is unlikely to have been as effective as the contracting program given the cross country evidence and the large program effects we saw given the modest additional outlay. In addition, financial controls within the government of Cambodia are weak.

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.

Overall, the contracting project was very effective in improving service delivery in the project area. Loevinsohn & 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 the 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.

References

- Angrist, Joshua D., Imbens, Guido W., & Rubin, Donald B. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, **91**(434).
- Banerjee, Abhijit, Deaton, Angus, & Duflo, Esther. 2004. *Health Care Delivery in Rural Rajasthan*. Poverty Action Lab Paper 7.
- Bertrand, Marianne, Duflo, Esther, & Mullainathan, Sendhil. 2002. *How Much Should We Trust Differences-in-Differences Estimates?* NBER Working Paper 8841.
- Bhushan, Indu, Keller, Sheryl, & Schwartz, Brad. 2002. *Achieving the twin objectives of efficiency and equity: contracting health services in Cambodia*. Asian Development Bank ERD Policy Brief Series Number 6.
- Cambodia. 2002. *General population census of Cambodia, 1998 : final census results*. National Institute of Statistics.
- Cambodia. 2004. *The Medium-Term Expenditure Framework for Cambodia 2005-2007*.
- Chaudhury, Nazmul, Hammer, Jeffrey, Pruthi, Himani, & Ranganathan., Rupa. 2005. *Health, Wealth, and Openness: Changes in Child Morality During a Period of Market Liberalization in India*. Mimeo. The World Bank.
- Chaudhury, Nazmul, Hammer, Jeffrey, Kremer, Michael, Muralidharan, Karthik, & Rogers, F. Halsey. 2006. Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, **20**(1), 91–116.
- Conway, Tim. 2000. *Current Issues in Sector-Wide Approaches for Health Development. Cambodia Case Study*. World Health Organization.
- Das, Jishnu, & Sánchez-Páramo, Carolina. 2002. *Short but not Sweet: New Evidence on Short Duration Morbidities from India*. World Bank Policy Research Working Paper 2971.
- Donald, Stephen G., & Lang, Kevin. 2001. *Inference with Difference in Differences and Other Panel Data*.
- Filmer, Deon, & Pritchett, Lant. 1999. The impact of public spending on health: does money matter? *Social Science and Medicine*, **49**, 1309–1323.
- Greevy, Robert, Silber, Jeffrey, Cnaan, Avital, & Rosenbaum., Paul. 2004. Randomization Inference with Imperfect Compliance in the ACE-Inhibitor After Anthracycline Randomized Trial. *Journal of the American Statistical Association*, **99**(485), 7–15.
- Holmstrom, Bengt, & Milgrom, Paul. 1991. Multi-Task Principal-Agent Analysis:

- Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization*, **7**(0), 24–52.
- Imbens, Guido, & Angrist, Joshua. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.
- Keller, Sheryl, & Schwartz, Brad. 2001. *Final Evaluation Report: Contracting for Health Services Pilot Project*. Unpublished Asian Development Bank Report on Loan No. 1447-CAM.
- Kling, Jeffrey R., Liebman, Jeffrey B., Katz, Lawrence F., & Sanbonmatsu, Lisa. 2004. *Moving to Opportunity and Tranquility: Neighbourhood Effects on Adult Economic Self-Sufficiency and Health From a Randomized Housing Voucher Experiment*.
- Loevinsohn, Benjamin, & Harding, April. 2005. Buying Results? Contracting for Health Service Delivery in Developing Countries. *The Lancet*, **366**, 676–81.
- Logan, B.R., & Tamhane, A.C. 2003. *On O'Brien's OLS and GLS Tests for Multiple Endpoints*. Unpublished Ms.
- O'Brien, Peter C. 1984. Procedures for Comparing Samples with Multiple Endpoints. *Biometrics*, **40**(4), 1079–1087.
- Phadke, Anant. 1998. *Drug Supply and Use: Towards a Rational Policy in India*. Thousand Oaks, CA: Sage Publications.
- RACHA. 2000. *The Pathway to Child Health (Siem Reap, Pursat, Stung Treng, and Kampot)*.
- Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer.
- Schwartz, Brad, & Bhushan, Indu. 2004. *Reducing Inequity In The Provision Of Primary Health Care Services: Contracting In Cambodia*. Paper presented at World Bank Conference.
- Soeters, Robert, & Griffiths, Fred. 2003. Improving government health services through contract management: a case from Cambodia. *Health Policy and Planning*, **18**(1), 74–83.
- 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.
- Wooldridge, Jeffrey M. 2003. Cluster-Sample Methods in Applied Econometrics. *American Economic Review*, **93**(2), 133–138.
- World Bank. 2007. *World Development Indicators Online*.
- World Health Organization. 2001. *Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics and Health*. Geneva: World Health Organization.

Xu, Ke, Evans, David B, Kawabata, Kei, Zeramdini, Riadh, Klavus, Jan, & Murray, Christopher. 2003. Household Catastrophic Health Expenditure: A Multicountry Analysis. *The Lancet*, **362**(9378), 111–117.