Journal of Health Economics 31 (2012) 271–284

Contents lists available at SciVerse ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Financial incentives for maternal health: Impact of a national programme in Nepal

Timothy Powell-Jackson*, Kara Hanson

London School of Hygiene and Tropical Medicine, Health Economics and Systems Analysis Group, United Kingdom

ARTICLE INFO

Article history:
Received 3 November 2010
Received in revised form 25 August 2011
Accepted 28 October 2011
Available online 3 November 2011

JEL classification:
I10
I18

Keywords:
Impact evaluation
Financial incentives
Demand for health care
Maternal health
Nepal

ABSTRACT

Financial incentives are increasingly being advocated as an effective means to influence health-related behaviours. There is, however, limited evidence on whether they work in low-income countries, particularly when implemented at scale. This paper explores the impact of a national programme in Nepal that provides cash incentives to women conditional on them giving birth in a health facility. Using propensity score matching methods, we find that the programme had a positive, albeit modest, effect on the utilisation of maternity services. Women who had heard of the SDIP before childbirth were 4.2 percentage points (17 percent) more likely to deliver with a skilled attendant. The treatment effect is positively associated with the size of the financial package offered by the programme and the quality of care in facilities. Despite the positive effect on those exposed to the SDIP, low coverage of the programme suggests that few women actually benefited in the first few years.

1. Introduction

Access to priority health services in low-income countries remains vastly inadequate. Nowhere is this more obvious than in maternal health. According to a widely cited paper by Campbell and Graham (2006), a strategy in which women give birth in primary care institutions with effective referral is key to improving maternal health. Yet, improvements in the coverage of professional care at childbirth has stagnated in Sub-Saharan Africa and South Asia over the past decade, in part, because the provision of and the reluctance to use maternity services are inextricably linked to deep-rooted issues such as the state of the health system and the place of women in society (Koblinsky et al., 2006). Because professional care at childbirth is often used as a broader proxy for the state of a health system (Rohde et al., 2008), these trends raise concerns beyond maternal health.

In response financial incentives have been increasingly advocated as an effective means to change health-related behaviours and improve health outcomes (NORAD, 2007). If households lack the financial resources, heavily discount the future or lack information on the benefits of health care to make optimal care seeking choices, financial incentives can increase demand for health care. Financial incentives are the key feature of various programmes that have become popular in recent years, including conditional cash transfers, vouchers and one-off cash payments.

Financial incentives provide an immediate reward to individuals for behaviour that leads to health gains, and have been used to target a range of health-related behaviours. Recent enthusiasm for their use in low and middle income countries is supported by evidence showing that payments aimed at initiating take up of preventive health interventions can be effective (Lagarde et al., 2007). However, there is also limited evidence of perverse effects.[1] Financial incentives have been used in a positive sense, to encourage uptake of health technologies and attendance at health clinics (Fiszbein et al., 2008). More controversially perhaps, they have been used to encourage individuals to refrain from certain behaviours, such as contracting sexually transmitted diseases (Jack, 2008).

In this paper we explore the effect of a national financial incentive programme for maternal health in Nepal. Despite a recent improvement in maternal mortality in the country, utilisation of maternity services has remained unacceptably low (Pradhan et al.,

* Corresponding author. Tel.: +44 020 7612 7887.
E-mail address: Timothy.Powell-Jackson@lshtm.ac.uk (T. Powell-Jackson).

[1] With a few notable exceptions, there is little evidence on undesirable effects of financial incentives in health. Conditional cash transfers in Honduras may have increased fertility because only pregnant women were eligible to benefit (Morris et al., 2004a). It is argued that children in Brazil may have been kept malnourished owing to a misperception that this would qualify the household for financial benefits (Morris et al., 2004b).

0167-6296/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2011.10.010

Table 1
Financial incentives offered by the SDIP.

Financial incentive	Eligibility criteria
1. Cash payment to women • 500 NRS ($7.8) in plains districts • 1000 NRS ($15.6) in hill districts • 1500 NRS ($23.4) in mountain districts	Woman delivered in a public health facility and had no more than two living children or an obstetric complication (as diagnosed by the health provider)
2. Provider incentive • 300 NRS ($4.7) for each delivery attended	Doctor, nurse, midwife, health assistant, auxiliary health worker or maternal and child health worker attended a delivery at the woman's home or in a public health facility
3. Free delivery care to women and facility reimbursed • 1000 NRS ($15.6) reimbursed to health facility	Woman comes from one of the 25 least developed districts and meets the eligibility criteria required for the financial incentive

1997; Government of Nepal, 2001, 2007). The Government of Nepal thus turned to the use of financial incentives. Introduced nationwide in July 2005, the Safe Delivery Incentive Programme (SDIP) provides: (i) a cash payment to women who give birth in a public health facility; (ii) exemption from user fees for those residing in the least developed one third of districts; and (iii) a financial incentive to health workers. The incentive to health workers is given for attendance at deliveries both in the health facility and at the home of the woman giving birth (Government of Nepal, 2005). As shown in Table 1, the amount of cash was designed to vary across the three main geographical regions of Nepal to reflect differences in the cost of accessing health services faced by households (Borghi et al., 2006a). The development of the SDIP and its rapid adoption was heavily influenced by a convergence of political interests and effective dissemination of research findings supporting the notion of financial incentives (Ensor et al., 2009). At the time, the coalition government was headed by the United Marxist Leninist party who, in their manifesto, had pledged support to advancing the status of women.

We focus on estimating the effect of the SDIP on women's use of health care services at childbirth. The cost of maternity care faced by households can be high, with the majority of expenditures made outside of the health facility (Borghi et al., 2006b). By reducing these costs, the SDIP is expected to lead to improved health seeking behaviour at childbirth. We estimate the magnitude of the effect on use of formal care and then seek to understand whether the benefits of the SDIP vary according to characteristics of the target population and the design of the programme. Variation in the package of financial benefits across regions, for example, provides an opportunity to explore whether the size of the incentive makes any difference to the impact of the programme.

Our empirical strategy relies on an unusual measure of treatment, namely the woman's knowledge of the SDIP prior to childbirth, and propensity score matching methods to estimate the causal impact of the programme. Identification rests on the strong assumption of conditional independence and, for this reason we explore a number of approaches to assess the robustness of the basic findings. Over our study period, implementation was characterised by lengthy delays in the disbursement of funds from the central level and hesitation on the part of the government to promote the programme using mass media (Powell-Jackson et al., 2009a). As we argue later, the extent of implementation must be given consideration when interpreting the findings.

The paper contributes to the growing literature on demand-side incentives in health. However, there is little rigorous evidence on whether financial incentives work in low-income countries, particularly when implemented at scale. The available evidence comes largely from middle-income Latin American countries, where health services are available and government financial systems relatively strong. This paper also informs the debate on the feasibility of implementing financial incentive programmes in resource-poor settings. While the notion of paying individuals to influence their behaviour is simple and intuitively appealing, our findings suggest that such interventions can be complex to implement (Oxman and Fretheim, 2008).

The paper is structured as follows: Section 2 considers the main theoretical mechanism underpinning the SDIP in formulating predictions as to its effect. Section 3 describes the methods, including our definition of treatment, the empirical strategy and the data used in the study. Section 4 presents the findings and Section 5 discusses the main implications and limitations of the study.

2. Theoretical considerations

We start with a conceptualisation of the pathways through which the SDIP can be expected to improve outcomes, in an effort to make explicit the assumptions that underpin the process (Weiss, 1998; White and Masset, 2007). This leads us to identify a number of steps that can be considered necessary if the programme is to lead to a change in health seeking behaviour. It also allows us to make the distinction between individual actions and government involvement in the implementation process and provides the basis with which we define our treatment group in the subsequent analysis.

At the individual level, the key steps include: households with a pregnant woman hear about the financial incentives offered by the SDIP; households perceive the promise of financial incentives as credible – that is, they expect to receive the benefits; women give birth with professional care in a health facility; and after giving birth women receive the demand-side incentives in a timely manner. Women's experience of the administration process is communicated to other families and this in turn, along with other factors, affects their expectation of whether they will receive the demand-side incentives in the future. The government's role in the implementation process is to promote the SDIP if target households are to know about the demand-side incentives on offer and to ensure funds are available in health facilities if women are to be paid on time. The latter requires a well functioning public financial management system since the funds must flow from the central treasury to each district health office and then to each health facility.

In formulating predictions about the effect of the SDIP on health care seeking, we emphasise the price mechanism as the primary causal pathway through which the demand-side incentives affect behaviour. A simple model of provider choice illustrates that an individual is essentially faced with a trade-off between health and non-health consumption (Gertler and Van der Gaag, 1990). The SDIP's demand-side incentive represents a subsidy on the price of care which will induce an increase in the demand for publicly provided maternity services. The increase in demand is the result of two effects: a substitution effect and an income effect. The former occurs because of a change in the relative prices of alternative providers, while the latter is the result of an increase in purchasing power.

The extent to which the demand-side incentives increase the use of formal care in the public sector (i.e. the price elasticity of demand) is an empirical question and the main focus of this paper. Since the SDIP operates only in the public sector, we anticipate a substitution away from home care and non-state providers. It can also be shown that poorer individuals have a lower price elasticity of demand for health care than wealthier individuals, as long as health is a normal good (Gertler et al., 1987). For the increase in demand to translate into utilisation, maternity services must be available and the quality of these services will determine whether

there are improvements in health outcomes. In other words, we expect the effect of the demand-side incentives to be greater where health services are available.

3. Methods

3.1. Defining treatment

Implicit in the archetypal evaluation problem with a binary treatment is a clear definition of the treatment status of each individual in the population of interest. The most common way of defining treatment uses enrolment status, eligibility status or geographical placement of the programme. In this study, all three were ruled out owing to the nature of the programme and the fact that it was launched nationwide from the outset.

Instead, the study design is informed by the model of the programme's causal pathway. The implementation process points towards knowledge about the SDIP prior to childbirth as a necessary condition for the programme to affect health seeking behaviour. A family's decision of where to seek care may be influenced by its ex-ante expectation of the price of care. It follows that the SDIP's demand-side incentives will only influence the health seeking behaviour of those families who have knowledge of the programme's benefits before childbirth. This definition of treatment implies that only the impact of the financial incentive and free delivery care – and not the health provider incentive – can be assessed. The impact estimates, therefore, do not reflect the possible supply-side influence of the provider incentive on utilisation, although we do recognise that the impact of the demand-side incentives may vary according to supply-side factors.

Household interviews were carefully conducted to ensure the knowledge of the women and the family decision-maker was captured to reflect the fact that few married women in Nepal make decisions regarding their own health care (Government of Nepal, 2007).[2] There were concerns that the SDIP might be confused with other health programmes and the use of ex post information on knowledge of the programme prior to childbirth might be susceptible to recall bias. Thorough pre-testing of the survey tool and discussions with respondents after each interview provided reassurance that the validity of the measure was not compromised by these two concerns. Specifically, women found the SDIP's offer of cash a highly distinctive feature. This meant that there was little risk of respondents confusing the SDIP with other health programmes since no other government programme offers cash to women. Furthermore, childbirth is a highly memorable event (whether it is positive or negative), allowing women to relate the timing of when they found out about the SDIP to the date of giving birth.[3] With treatment measured in this way, the survey tool aimed to collect information on factors that would be expected to influence jointly exposure to information about the SDIP and health seeking behaviour at childbirth.

3.2. Empirical strategy

We identify the impact of the SDIP using propensity score matching methods. The idea behind the approach is to select a comparison group of non-participants that is as similar as possible to the treatment group in its observed characteristics. Individuals in the comparison group are selected on the basis of their propensity score, given by $P(Z) = Pr(D = 1|Z)$ and $(0 < P(Z) < 1)$, where D is a treatment indicator and Z is a set of control variables unaffected by programme participation. Appropriate matching variables are those that jointly affect treatment status and the outcome. The propensity score gives the probability that an individual participates in the programme given the set of observed characteristics that jointly influence treatment status and outcomes. It is estimated by means of a probit model.

Two assumptions are required for the identification of the average treatment effect on the treated (ATT). First, the conditional independence assumption, or unconfoundedness, states that outcomes are independent of participation given the observables (i.e. $Y_0, Y_1 \perp D|Z$). It can be shown that the conditional independence assumption continues to hold conditional on $P(Z)$, such that outcomes are independent of participation given the propensity score (i.e. $Y_0, Y_1 \perp D|P(Z)$) (Rosenbaum and Rubin, 1983, 1984). Second, the common support assumption ensures that there are treated and untreated individuals with the same characteristics ($P = D = 1|Z < 1$).

When these conditions hold, the average treatment effect on the treated is identified non-parametrically by the mean conditional difference in the outcome over the common support, suitably weighted by the distribution of Z in the treatment group. There are a number of ways to construct the matched outcome that vary around how the set of neighbours, $C^0(Z_i)$, is defined and how the weights, W_{ij}, are chosen. We try three matching estimators and evaluate how well they balance the covariates.[4] With caliper matching, each treated observation is matched to the 10 neighbours nearest in terms of their propensity score (Cochran and Rubin, 1973). Untreated observations must fall within a maximum distance (caliper) of 0.01 in order to be matched. In kernel matching, the outcome of a treated individual is matched to the weighted outcomes of all untreated units, where the weight is in proportion to the closeness between the propensity score of the treated and untreated individuals (Heckman et al., 1998). The bandwidth is set at 0.03. Finally, Mahalanobis-metric matching combines the matching variables into a distance measure and then matches based on the resulting scalar. The propensity score is included in the set of matching variables.

Although we use a rich set of data from a survey designed specifically to evaluate the SDIP and capture information on some of these 'unobservables,' conditional independence remains a strong assumption and the basic results are unlikely on their own to provide fully convincing evidence of a causal effect. We pursue three strategies to mitigate concerns about bias due to potential correlation between exposure to the SDIP and unobserved factors affecting health seeking behaviour. First, we analyse the impact of the SDIP by the type of provider to explore whether there is any evidence of a substitution effect. If the ATT estimates are subject to omitted variable bias, this bias can be expected to work in the same direction for state and non-state health providers, given that they are

[2] We refer to the *woman's* knowledge of the SDIP throughout the paper, but this should always be interpreted as the family's knowledge of the SDIP.

[3] The most common misreporting is expected to be women stating that they knew about the SDIP during pregnancy when in fact they found out subsequently. This would downward bias impact estimates since these women, incorrectly classified as treated individuals, were not exposed to information about the SDIP and are thus less likely to have delivered in a health facility.

[4] The percent of treated observations lost due to common support and various tests of matching quality provide the basis with which to evaluate the matching procedures. Smith and Todd (2005) suggest testing for differences in the covariates between the treated and the non-treated group. For each variable, the standardised percentage bias – the difference of the sample means in the treated and non-treated groups as a percentage of the square root of the average of the sample variances in each group (Rosenbaum and Rubin, 1985) – is calculated before and after matching. A two-sample t-test can then be used to check if there are significant differences between the means before and after matching. In addition, a likelihood-ratio test of joint significance of covariates before and after matching can provide tests of matching quality.

the closest substitutes. Impact estimates that are upward biased are likely to hide all evidence of a substitution effect between the different types of provider.

A second strategy is akin to providing evidence on a dose–response relationship. It builds on the insight that knowledge of the SDIP is necessary but insufficient if the programme is to have an impact. There must also be a strong expectation that the woman will receive the cash incentive after childbirth. We redefine treatment to create two new treatment groups. T^{CT} is made up of women who knew about the SDIP prior to childbirth and expected to receive the cash. T^{PT} consists of women who knew about the SDIP but did not expect to receive the cash incentive. Expectations of receiving the cash are measured by whether the woman knew of anyone else who had received the cash incentive of the SDIP.[5] Propensity score matching is applied to each of these treatment groups in separate analyses using the same comparison group as previously. Given that we expect the treatment effect in the analysis with T^{PT} as the treatment group to be zero coupled with the fact that women in the comparison group and treatment group have very different characteristics (and different biases), any omitted variable bias should be apparent through a difference in the mean utilisation of maternity services between the two groups (Imbens and Wooldridge, 2009). In this sense, a finding of no treatment effect in the analysis with T^{PT} as the treatment group, and a large treatment effect in the analysis with T^{CT} as the treatment group provides evidence in support of the assumption of conditional independence.

Third, we control for unobservables using an instrumental variable strategy to explore whether omitted variable bias plagues the propensity score estimates. A promising candidate to instrument knowledge of the SDIP during pregnancy is the frequency with which women in the sample listen to the radio. We know from the survey that some women heard about the SDIP from the radio and indeed district health offices reported separately in qualitative interviews that they used the radio to promote the programme. The radio was not used, however, to promote institutional delivery care more generally.[6] Thus, to the best of our knowledge, the only pathway through which the radio influences a woman's place of delivery is through its promotion of the SDIP. The two-stage least squares estimate of impact should be interpreted as a local average treatment effect, given that the assumption of homogeneous treatment effects is particularly unrealistic in this instance (Imbens and Angrist, 1994). The treatment effect, in other words, applies only to the sample of women who found out about the SDIP through the radio.

A final analysis explores heterogeneity in the impact of the SDIP (Djebbari and Smith, 2008). The effect of the programme is anticipated to vary along a number of dimensions, including household wealth, the level of benefits offered in a particular district and the availability or quality of care in the public sector. We analyse heterogeneity in impacts across our observables using subgroup analysis, with a view to disentangling which factors are associated with variation in the ATT estimates. Variation in impact by wealth, for example, may be due to a correlation between a household's wealth and the level of benefits provided in its district when it is in fact the latter which drives the relationship (Wagstaff et al., 2009). In a linear probability model, we regress the utilisation outcome on our treatment variable, the covariates and a set of interactions between treatment and the covariates. The regression is weighted by the kernel weight (Leuven and Sianesi, 2003).[7]

We include in the analysis of heterogeneous impacts a measure of the availability of maternity care for each type of government health facility.[8] We consider this measure a proxy for quality of care. In a survey of health providers conducted in parallel with the household survey, we asked each health facility whether it had provided a number of services that are considered critical components of maternal health care (UNICEF, 1997). Specifically, we asked if the health facility had (in the last three months): (i) administered parenteral antibiotics; (ii) administered parenteral oxytocics; (iii) administered anticonvulsants; (iv) performed manual removal of placenta; (v) performed manual removal of retained products; (vi) performed assisted vaginal delivery; (vii) given a blood transfusion; (viii) performed a caesarean section; (ix) referred a woman by ambulance; (x) provided 24 h delivery care services. The measure we use is the number of procedures that the health facility was able to carry out (0–10). The variables are measured at the district level, taking the mean value if there is more than one facility of that type in the district.

3.3. Data

The data used in the analysis come from a survey of women that was conducted two and a half years after the start of the SDIP. Women who had given birth in the previous three years were interviewed across six districts of Nepal. Two districts were randomly selected from each of three ecological regions of Nepal to ensure we had variation in the level of benefits offered by the SDIP across the districts. The sample was selected in two stages: in the first stage 180 primary sampling units (villages in rural areas or urban wards) were selected using systematic sampling with probability proportional to size, and in the second stage an average of 30 women in each village were chosen randomly.

Each observation is a delivery and, given the three year recall period, it is possible for a woman to have had more than one delivery. The recall period extended five months before the start of the SDIP. We dropped two observations for lack of data. The dataset contains complete data on 5903 deliveries, of which 420 took place before the start of the SDIP. In the sample, women were unaware of the SDIP prior to childbirth in three-quarters of cases. These women make up our comparison group.

The impact of the SDIP is assessed on a set of binary utilisation outcomes that refer to the place of delivery, the type of attendant present at the delivery and the type of procedure,[9] if carried out during the delivery. Utilisation of professional delivery care services in the comparison group is low (Table 2). Deliveries in a health facility account for 16 percent of all deliveries and the public sector is the dominant provider. Just over 10 percent of deliveries in the comparison group take place in a government health facility, while non-governmental (not-for-profit) and

[5] The rationale for this measure is that a person's perception of the administration of the programme is particularly sensitive to the experiences of others in the community. Clearly, this indicator is a crude measure of expectations given that it is binary and fails to capture the full range of values from a probability. Nevertheless, it seems reasonable that expectations of receiving the cash incentive will be higher among those who knew of someone who had received the money than those who had never heard of anyone.

[6] The evaluation of the SDIP used research teams drawn from each of the study districts. These individuals had a deep knowledge of the local area and were able to confirm that FM radio was used solely to promote the SDIP.

[7] As noted by Wagstaff et al. (2009), the standard errors from the regression are smaller than those generated by psmatch2, which means in our analysis of heterogeneous effects we report t-statistics that are somewhat higher than warranted.

[8] Through its supply-side incentives, the SDIP may have improved the availability of maternity services and thereby increased use of maternal health care. This possibility, however, should not threaten the internal validity of our main findings which isolate the demand-side effect of the programme.

[9] An assisted delivery refers to the use of forceps or a ventouse that are attached to the baby's head.

Table 2
Descriptive statistics, by treatment group.

Variable	Treated		Comparison	
	Mean	Std deviation	Mean	Std deviation
Outcomes				
Health facility	0.263	0.440	0.155	0.362
Government health facility	0.210	0.407	0.106	0.307
NGO hospital	0.024	0.154	0.031	0.175
Private health facility	0.029	0.168	0.018	0.134
Doctor, nurse or midwife in attendance	0.293	0.455	0.182	0.386
Any health worker in attendance	0.351	0.478	0.225	0.418
Caesarean section	0.047	0.212	0.025	0.155
Caesarean section or assisted	0.099	0.299	0.048	0.214
Covariates				
Age of woman	25.379	5.285	26.622	5.990
Log of wealth	3.062	0.658	2.911	0.676
No education (reference)				
Primary education	0.135	0.342	0.119	0.323
Secondary education	0.271	0.445	0.195	0.396
Higher education	0.189	0.391	0.080	0.272
No work (reference)				
Agriculture work	0.684	0.465	0.701	0.458
Salaried work	0.030	0.169	0.012	0.109
Small business	0.066	0.248	0.046	0.209
Waged work	0.063	0.243	0.084	0.278
Other work	0.003	0.058	0.002	0.045
Brahmin and Chhetri (reference)				
Terai and Madeshi	0.063	0.243	0.079	0.270
Dalit	0.118	0.323	0.141	0.348
Newar	0.043	0.203	0.016	0.125
Janajati	0.173	0.379	0.204	0.403
Muslim	0.005	0.068	0.028	0.165
Other castes	0.075	0.263	0.129	0.335
Walk to facility < 1 h (reference)				
Walk to facility 1 h < 4 h	0.613	0.487	0.572	0.495
Walk to facility 4 h < 1 day	0.115	0.319	0.184	0.388
Walk to facility > 1 day	0.031	0.173	0.106	0.307
Urban dwelling	0.120	0.325	0.084	0.277
Previous delivery during SDIP	0.091	0.287	0.060	0.238
Active FCHV	0.147	0.128	0.076	0.098
Women's groups	0.337	0.473	0.274	0.446
Morang (reference)				
Sankhuwasabha	0.238	0.426	0.140	0.347
Myagdi	0.163	0.370	0.142	0.350
Rupandehi	0.109	0.312	0.160	0.367
Jumla	0.217	0.412	0.185	0.389
Achham	0.129	0.335	0.189	0.391
Year	1.947	0.752	1.617	0.872
Observations	1489		4416	

Note: FCHV = female community health volunteer.

private (for-profit) health facilities account for 3 percent and 2 percent of all deliveries respectively. Almost a fifth of women in the comparison group give birth with a doctor, nurse or midwife in attendance, the indicator that corresponds with the standard international definition of a skilled birth attendant. The rate of caesarean section in the study area is low, with only 3 percent of women having surgery at childbirth. While the survey was not designed to be representative of the entire country, the comparison group estimates correspond very closely with national estimates, which mostly covered the period before the SDIP (Government of Nepal, 2007).

The decision of what covariates to include is informed by the various sources of information from which women found out about the SDIP. Table 2 provides summary statistics of the covariates used in the propensity score matching. Most, but not all, of the covariates are self-explanatory. A wealth index is constructed from information on the ownership of assets using principal components analysis and re-scaled to give values between 1 and 100 (Filmer and Pritchett, 1998). We take the log of the re-scaled index. We include two village level variables. The level of activity of the female community health volunteer is an index that ranges between possible values of 0–1. It provides a measure of how active the female community health volunteer in each village is in disseminating information about the SDIP to the community. It is calculated as the proportion of women in a village who found out about the SDIP through the female community health volunteer. Each observation takes the mean value for the village. The women's group variable indicates whether the village in which the woman lives has women's group meetings. The year variable takes the value of the fiscal year in which the delivery took place with possible values ranging from 1 to 4, and 2 representing the first year of the SDIP. It captures unobservables whose variation over time influence both programme participation and outcomes.

Based on these covariates, the two groups of women appear to be different. The comparison group is older, less wealthy, less educated, more likely to be from marginalised castes, and lives further from its nearest health facility than the treatment group.

Table 3
Covariate balancing indicators before and after matching.

Matching estimator	N_1 Before	N_0 Before	Probit pseudo R^2 Before	Probit pseudo R^2 After	$p > \chi^2$ After	Mean bias (%) Before	Mean bias (%) After	Lost to common support (%) After
Kernel matching	1489	4416	0.177	0.002	1.000	17.18	1.74	0.13
Malahanobis matching	1489	4416	0.177	0.021	0.000	17.18	3.46	0.00
Nearest neighbour matching	1489	4416	0.177	0.002	1.000	17.18	1.88	0.81

Note: The pseudo R^2 is from the probit estimation of the conditional treatment probability and gives a measure of how well the regressors explain variation treatment. The p value of the likelihood ratio test after matching tests the hypothesis that the regressors are jointly significant. The mean bias is the mean absolute standardised bias, as defined in Section 3.2. Lost to common support gives the percentage of treated observations that fall outside the boundaries.

3.4. Propensity score estimation and balancing

The results of the probit model are shown in Table A1 in the Appendix. While the probit model is primarily used as a statistical tool to estimate the propensity score, it can also shed light on inequalities in implementation. The pseudo R^2 of the probit model is 0.177 and the coefficients indicate that a handful of covariates are important predicators of treatment status. Younger, more educated, and wealthier women are more likely to have knowledge of the SDIP during pregnancy. Women belonging to Janajati or Muslim castes are less likely than the reference group (Brahmin/Chhetri) to have knowledge of the programme, while those living closer to a health facility are more likely to have heard of the programme, as are women living in villages with active female community health volunteers. If a woman gave birth to a previous baby after the start of the SDIP, she has a higher chance of knowing about the programme. Women from the districts of Sankhuwasabha, Myagdi, Rupandehi and Achham are more likely than those in Morang to know about the SDIP while they were pregnant. Lastly, as the programme has matured, exposure to information on the programme has increased, as indicated by the effect of the year variable.

The purpose of the matching exercise is to balance the covariates such that the bias on observables is reduced as much as possible. Overall measures of imbalance in the covariates before and after matching are reported in Table 3. The pseudo R^2 from the probit model estimated on the trimmed sample using the weights generated from all three matching exercises is appreciably lower than the value from the model on the original unweighted sample. Kernel matching and nearest neighbour matching perform best, reducing the pseudo R^2 value from 0.177 to 0.002 respectively.

The hypothesis of the joint insignificance of all the regressors before and after matching cannot be rejected in the case of the kernel matching and the nearest neighbour matching, as indicated by the p-values of the likelihood-ratio test. Before matching, the average standardised bias is 17.2 percent. All three matching methods reduce the bias dramatically, although some more than others. Kernel matching reduces the average bias the most to 1.7 percent, while caliper matching reduces the average bias to 1.9 percent. Mahalanobis matching performs less well, reducing the average bias to 3.5 percent. Further evidence on the quality of matching is provided in Table A2 in the Appendix.

The percentage of observations lost to the common support restrictions is negligible indicating that none of the matching methods appear to pose a problem in this regard. The kernel, Malahanobis and caliper matching procedures lose two (0.13%), zero and twelve (0.81%) treated observations respectively. Fig. 1 shows the histogram of the propensity scores before matching for those in the treatment group and those in the comparison group. The region of common support is substantial despite there being a right-skewed distribution in the case of the untreated deliveries.

The common support and balancing test results suggest that bias associated with differences in the observables between the treated and untreated groups can be almost completely eliminated without the need to discard a large number of observations from the sample due to a lack of common support. Since the region of common support includes almost the entire sample, the estimated treatment effect barely has to be redefined, essentially allowing the ATT to be recovered. While there is little to choose between the kernel and the nearest neighbour matching procedures, the former is preferred on the basis that it loses the least observations to common support while achieving the greatest reduction in bias. The next section reports estimates of programme impact using kernel matching.

4. Results

4.1. Effectiveness of implementation

To provide some context behind the impact results, we first examine how well the programme was implemented. Two measures are particularly revealing. The first concerns awareness of the SDIP among the target population. Just under a quarter (24.3 percent) of women in our sample had knowledge of the SDIP prior to childbirth. This estimate implies that three-quarters of women were not reached by the programme and their health seeking behaviour could not plausibly have been influenced by the offer of the financial incentive. In Fig. 2, the concentration curve of knowledge of the SDIP shows evidence of some inequality, particularly in the poorest wealth quintile.

The second measure of implementation concerns receipt of the financial incentive. Only 26.5 percent of women who gave birth in a government health facility were given the financial incentive or an equivalent deduction on their bill. There were clearly considerable problems in paying women. The vast majority of women who were meant to receive the financial incentive were never paid. Inequality in the receipt of the financial incentive across the entire sample

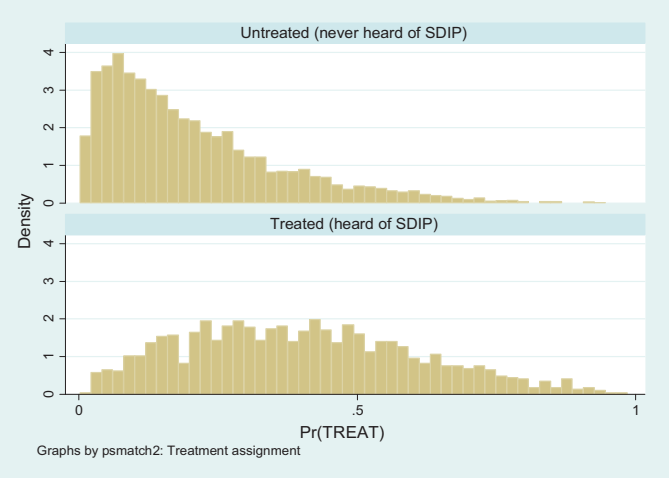

Fig. 1. Histogram of propensity scores for the treated group and comparison group.

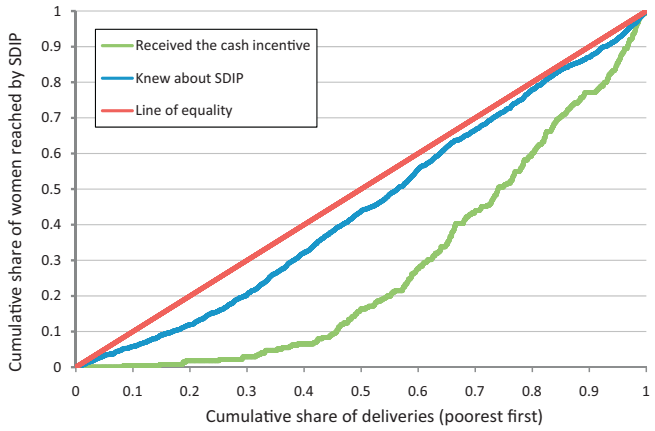

Fig. 2. Concentration curves for knowledge of the SDIP prior to childbirth and receipt of the financial incentive.

of women gives an indication of the benefit incidence. As the concentration curve in Fig. 2 shows, the financial incentive is captured disproportionately by relatively wealthier households. While this largely reflects inequality in the use of government health services, it does serve to show that without any targeting such an intervention effectively subsidises the care of wealthier households who use services.

Findings from a process evaluation, conducted in parallel with this study shed light on why implementation in the first few years of the SDIP was so incomplete (Powell-Jackson et al., 2009a). The key implementation challenges included bureaucratic delays in the disbursement of funds and difficulties in communicating the policy to the public and health providers. There were lengthy delays in the transfer of funds from the central government to the districts. Districts on average received funds earmarked for the SDIP 283 days late in the first fiscal year, and 147 days late in the second fiscal year, with much of the delay caused by the late transfer of funds from the international donor to the central government. Because the central government was then wary of raising expectations among the public without having the funds to pay beneficiaries, programme managers chose not to publicise the SDIP with a national media campaign.

These two problems at the central level provide a convincing explanation for the broad findings presented above. It should be noted, however, that some districts were able to cope remarkably well with these challenges, providing an explanation for why uptake of the programme varied considerably between districts. Actions at the district level appeared to be have been influenced by the pressure to meet local needs, as well individual perceptions and acceptance of the programme among district implementers.

4.2. Main impact results

Table 4 shows estimates of the ATT for each utilisation outcome with the corresponding *t*-statistic, the percentage change relative to the mean of the untreated and the 95% confidence interval around the relative change. In column 1, estimates before matching indicate that there are substantial differences in utilisation between the treated and untreated groups. However, all this serves to show is that there is likely to be selection bias on the observables.

The matched estimates indicate that the SDIP, for those it reached, had a statistically significant effect on the probability of women delivering in a health facility. Institutional deliveries for those treated increased by 4.0 percentage points (relative change 18%; CI: 5%, 31%) as a result of the SDIP. Among those reached by the SDIP, impact estimates show that the programme increased both deliveries with a skilled birth attendant and deliveries with any professional health worker by a statistically significant 4.2 percentage points (relative change 17%; CI: 4%, 29%) and 5.2 percentage points (relative change 17%; CI: 6%, 28%), respectively.

The programme also had a positive effect of 1.2 percentage points (relative change 36%; CI: −3%, 74%) on the caesarean section rate (significant at the 10 percent level), and a positive impact of 1.9 percentage points (relative change 24%; CI: 0%, 48%) on caesarean sections and assisted deliveries combined. The results are robust to the matching procedure used (see Table A3 in the Appendix). There is consistency in the statistical significance and magnitude of impact estimates across the different outcomes. Programme impacts tend to be greatest with the nearest neighbour matching.

The findings provide encouraging signs that the SDIP increased use of formal maternal health care. The positive effect of the SDIP on the caesarean section rate is more difficult to interpret because there is no way of knowing whether the additional procedures were medically required. But, given how low the caesarean section rate is in Nepal and the fact that there are no obvious incentives to the provider for performing more (unnecessary) caesarean sections, this result is likely to reflect an improvement in welfare.

4.3. Impact by type of provider

Any substitution effect from the reduction in the price of government maternity services should become apparent when we analyse

Table 4
Impact of the SDIP on health seeking behaviour.

	Before matching		After matching			
	Mean difference	*t*-Stat	ATT	*t*-Stat	ATT as % of mean of untreated	95% confidence interval
Place of delivery						
Health facility	0.107	9.33	0.040***	2.70	17.8	(5.1; 31.1)
Type of attendant						
Doctor, nurse or midwife (SBA)	0.110	9.11	0.042***	2.72	16.6	(4.1; 29.1)
Any professional health worker	0.126	9.70	0.052***	3.17	17.2	(6.1; 28.1)
Procedure at delivery						
Caesarean section	0.022	4.35	0.012*	1.83	35.5	(−3.1; 74.1)
Caesarean section or assisted	0.051	7.15	0.019**	2.02	23.9	(0.1; 48.1)

Note: Sample includes 5901 deliveries. *t*-Statistics are based on standard errors provided by psmatch2. Kernel matching with a bandwidth of 0.03 is used to compute impact estimates. A treated observation whose propensity score is greater than the largest of the propensity scores of the untreated is dropped. 95% confidence intervals are around the relative percentage impact. ATT = average treatment effect on the treated. SBA = skilled birth attendant.

* Impact parameters are significant at 10%.
** Impact parameters are significant at 5%.
*** Impact parameters are significant at 1%.

Table 5
Impact of the SDIP on health seeking behaviour by type of health facility.

	ATT	*t*-Stat	ATT as % of mean of untreated	95% confidence interval
Health facility	0.040***	2.70	17.8	(5.1; 31.1)
Government health facility	0.043***	3.28	25.8	(10.1; 42.1)
NGO hospital	−0.011*	−1.86	−31.4	(−65.1; 2.1)
Private health facility	0.008	1.40	36.3	(−16.1; 88.1)

Note: Sample includes 5901 deliveries. *t*-Statistics are based on standard errors provided by psmatch2. Kernel matching with a bandwidth of 0.03 is used to compute impact estimates. A treated observation whose propensity score is greater than the largest of the propensity scores of the untreated is dropped. 95% confidence intervals are around the relative percentage impact. ATT = average treatment effect on the treated. NGO = non-governmental organisation.

* Impact parameters are significant at 10%.
** Impact parameters are significant at 5%.
*** Impact parameters are significant at 1%.

the impact of the SDIP on utilisation by type of provider. Table 5 presents the ATT estimates for various types of health facility. The impact of 4.3 percentage points (relative change 26%; CI: 10%, 42%) in the government sector is offset by a negative effect of 1.1 percentage points (relative change −31%; CI: −65%, 2%) on utilisation of NGO hospitals. These results suggest that the SDIP encouraged women to switch from NGO hospitals to government health facilities. Reassuringly, this substitution effect did not account for the entire increase in the utilisation of government providers and, as noted in Table 4, there remains a statistically significant net positive impact on institutional delivery care.

Intuitively, this finding makes sense. Government and NGO health providers are close substitutes and it seems plausible that a subsidy targeting only the public sector encourages some women to switch from the NGO sector. The private for-profit sector, on the other hand, caters to only the wealthiest who are less likely to be incentivised by the offer of the SDIP's cash.

This finding also serves to strengthen the robustness of the basic results. Since health facilities in the public and NGO sectors are close substitutes, any bias on unobservables can be expected to operate in the same direction for both types of provider. Thus, if omitted variable bias was driving the positive impact on utilisation of government health facilities, it seems unlikely that there would be any evidence of a negative effect on deliveries in NGO hospitals.

4.4. Credibility of the SDIP's promise of cash

In the following analysis, we assume that it is not sufficient to have knowledge of the SDIP for the programme to influence health seeking behaviour. There must also be a strong expectation that the cash incentive will be given to the woman after childbirth. Table 6 reports the findings from the two analyses in which the treatment group is defined differently.[10] In column 1, the ATT estimates suggest that the SDIP had a large and significant impact on women who knew about the SDIP and had heard of someone receiving the financial incentive. These women increased utilisation of government maternity services by 9.6 percentage points (relative change 53%; CI: 31%, 74%) because of the SDIP.

In column 5, the results suggest that knowing about the SDIP while having little expectation of receiving the cash incentive has no effect on utilisation of maternity services. Knowledge on its own does not appear sufficient to change behaviour. The impact estimates on the key outcomes are small in magnitude and highly insignificant. A woman in this treatment group is 0.3 percentage points (relative change −2%; CI: −21%, 17%) less likely to deliver in a government health facility as a result of the SDIP and 0.2 percentage points (relative change −0.9%; CI: −15%, 13%) less likely to deliver with skilled birth attendance. In columns 4 and 8 we can see the 95% confidence intervals of the treatment effects from the two sets of results do not overlap for these outcomes.

The results from this analysis are in line with our expectation and can be interpreted as providing evidence in support of the robustness of the basic findings. The finding of no treatment effect and the fact that the pseudo treatment group (T^{PT}) and the comparison group are very different from each other (and therefore likely to have different biases) suggest that the conditional independence assumption is plausible. If conditional independence did not hold, the ATT estimates from the analysis of the pseudo treatment group (T^{PT}) would most likely be biased upwards, showing a spurious positive impact on utilisation.

The magnitude of the effect of the SDIP depends not only the quality of implementation, as suggested by these results, but also on prior expectations of receiving the financial incentive.[11] If individuals at the outset have an overly optimistic view of the government and over time are made to revise their expectation of receiving the financial incentive down, there may be longer term implications for the success of the SDIP and indeed other government policies. We can examine this possibility by estimating the effect of the SDIP over time, interacting six-monthly splines with the treatment indicator. Although the coefficients on these interactions are not significant, the pattern suggests that the effect of the SDIP was greatest in the first six months of the programme and diminished subsequently (result not shown).[12]

4.5. Instrumental variable estimates

To further address concerns over endogeneity, we instrument knowledge of the SDIP prior to childbirth using the frequency of radio listenership. Table 7 reports the impact estimate on utilisation of government maternity services and for the sake of convenience produces ordinary least squares (columns 1) and propensity score matching (columns 2) results.

Column 3 confirms that the instruments are correlated with knowledge of the SDIP prior to childbirth (*F*-statistic is 8.76). The Sargan test of over-identification passes comfortably (Sargan, 1958). The impact estimates using two-stage least squares are, if anything, greater in absolute size than those obtained using propensity score matching and ordinary least squares. These results suggest the propensity score matching estimates may not be subject to upward bias owing to omitted variables. The two-stage least squares estimates of the effect of the SDIP are not, however,

[10] In terms of the balancing of the covariates, we hasten to add that the kernel matching estimator performs well in both matching exercises.

[11] We thank an anonymous for pointing this out to us.

[12] This finding is also consistent with the results presented in Powell-Jackson et al. (2009b), which uses longitudinal data from a community surveillance system in one district to show that use of maternity care "jumped" at the start of the SDIP despite the fact that disbursement of the financial incentive in the first few months of the programme was particularly patchy.

Table 6
Credibility of the SDIP's promise of the cash incentive.

	T^{CT} as treatment group				T^{PT} as treatment group			
	ATT	*t*-Stat	ATT as % of mean of untreated	95% confidence interval	ATT	*t*-Stat	ATT as % of mean of untreated	95% confidence interval
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Place of delivery								
Health facility	0.094***	4.36	38.3	(20.7; 55.8)	−0.008	−0.47	−3.8	(−20.1; 12.1)
Government health facility	0.096***	4.87	52.6	(31.0; 74.2)	−0.003	−0.21	−2.0	(−21.1; 17.1)
NGO hospital	−0.015*	−1.85	−38.5	(−80.2; 3.1)	−0.008	−1.19	−24.6	(−66.1; 17.1)
Private health facility	0.012	1.48	52.6	(−18.5; 123.7)	0.003	0.55	16.8	(−44.1; 78.1)
Type of attendant								
Doctor, nurse or midwife (SBA)	0.090***	4.06	33.0	(16.7; 49.2)	−0.002	−0.13	−0.9	(−15.1; 13.1)
Any professional health worker	0.103***	4.45	31.7	(17.5; 46.0)	0.006	0.33	2.2	(−11.1; 15.1)
Procedure at delivery								
Caesarean section	0.023**	2.16	57.3	(4.2; 110.3)	0.002	0.27	6.1	(−39.1; 51.1)
Caesarean section or assisted	0.045***	3.04	49.9	(17.1; 82.8)	−0.001	−0.10	−1.5	(−32.1; 29.1)

Note: Sample includes 5093 deliveries in the comparison of treated and control observations. Sample includes 5223 deliveries in the comparison of pseudo control and control observations. *t*-Statistics are based on standard errors provided by psmatch2. Kernel matching with a bandwidth of 0.03 is used to compute impact estimates. A treated observation whose propensity score is greater than the largest of the propensity scores of the untreated is dropped. 95% confidence intervals are around the relative percentage impact. ATT = average treatment effect on the treated. SBA = skilled birth attendant. NGO = non-governmental organisation.

* Impact parameters are significant at 10%.
** Impact parameters are significant at 5%.
*** Impact parameters are significant at 1%.

Table 7
Dealing with endogeneity concerns.

Method	OLS	PSM	2SLS
	(1)	(2)	(3)
Knowledge of SDIP	0.038***	0.043***	0.046
	(0.011)	(0.013)	(0.153)
Controls	Yes	Yes	Yes
F test on instruments in first stage			8.76
Over-identification test *p* value			0.64
Adjusted R^2	0.21	–	0.22
Observations	5903	5903	5903

Note: Standard errors are reported in parentheses (robust standard errors in columns 1and 3). Controls include age, wealth (fourth degree polynomial), education, occupation of household head, caste, distance to nearest health facility, urban residence, previous delivery during SDIP, active FCHV index, availability of women's groups, and district fixed effects. The three instruments used in column 3 are: (i) listens to radio daily; (ii) listens to radio at least once a week; (iii) listens to radio less than once a week. The over-identification test is due to Sargan (1958).

* Impact parameters are significant at 10%.
** Impact parameters are significant at 5%.
*** Impact parameters are significant at 1%.

close to being statistically significant. The loss of efficiency and subsequently large standard errors are probably the result of two problems. First, there is multicollinearity between the instruments and several of the covariates, particularly education, which means the variation left over in the error term in the first stage is vastly reduced (Angrist and Pischke, 2009). Second, the fact that the dependent variable is binary increases the risk of a type II error through lack of statistical power.

4.6. Heterogeneity in impacts

We analyse variation in the impact of the SDIP on utilisation of government delivery care services. For various sets of variables, we test the null hypothesis that the coefficients of the interaction terms are equal to zero. Table 8 shows the results from three specifications of a linear probability model. In the first model, we include the full set of covariates used in the estimation of the basic results, but rather than having district dummy variables we instead include region dummy variables to capture, inter alia, differences in the package of financial benefits. We are able to reject the hypothesis that the coefficients on all the interaction terms are equal to zero, implying that the treatment effect is heterogeneous. The effect of wealth on the ATT estimate is negative but the coefficient is not significant at any level. Inclusion of wealth quintile dummy variables does not alter this result. The impact of the programme does, however, vary significantly by region, as shown by the rejection of the null hypothesis that the region dummy variables are equal to zero.

In the second model, the region dummy variables are replaced by the size of the SDIP's financial benefit as a share of the total cost of delivery to the household. The coefficient on the interaction between this variable and treatment is positive and significant, suggesting that the higher the subsidy relative to the cost of care, the greater the effect of the SDIP on utilisation. In the third model, we introduce the proxy measure of quality of care for each type of government health provider. The effect of the size of the SDIP's financial package is now even greater. We find that the quality of care at hospitals and primary health care centres is positively associated with the impact of the SDIP on utilisation of government delivery care services. The quality of care in hospitals modifies the effect the most, which seems plausible given that the majority of women who use the public sector deliver in hospitals. While these results are far from conclusive, they do provide some support to the common sense notion that demand-side incentives are more effective when the supply of health services are reliably in place.

5. Discussion

This paper finds that the SDIP had a positive impact on the utilisation of maternity services. Women in the treated group were 4.3 percentage points (26 percent) more likely to deliver in a public health facility, 4.2 percentage points (17 percent) more likely to deliver with a skilled birth attendant and 1.2 percentage points (36 percent) more likely to have a caesarean section. There was evidence, albeit weaker, that the SDIP encouraged women to substitute from NGO hospitals to give birth in government health providers. The SDIP, in other words, may have led to some crowding out of the private not-for-profit sector.

While there is considerable enthusiasm for the use of demand-side incentives in resource-poor settings, it has been noted that they may be inappropriate if health infrastructure is inadequate

Table 8
Variation in the impact of the SDIP on utilisation of government maternity services.

Variable	(1)		(2)		(3)	
	Coefficient	Std error	Coefficient	Std error	Coefficient	Std error
Treatment	0.131	0.172	−0.087	0.134	−1.204**	0.468
Treatment × age of woman	0.002	0.002	0.003	0.002	0.002	0.002
Treatment × log of wealth	−0.018	0.036	0.021	0.028	−0.040	0.036
Treatment × primary education	0.025	0.033	0.023	0.033	0.015	0.035
Treatment × secondary education	−0.035	0.036	−0.045	0.036	−0.046	0.038
Treatment × higher education	−0.071	0.059	−0.079	0.059	−0.074	0.064
Treatment × woman works agriculture	−0.069	0.045	−0.067	0.045	−0.053	0.046
Treatment × woman is salaried worker	0.016	0.107	0.008	0.107	0.024	0.107
Treatment × woman has small business	−0.015	0.075	−0.023	0.075	−0.001	0.075
Treatment × woman is a waged worker	0.005	0.071	0.021	0.071	0.012	0.073
Treatment × woman has other work	−0.508**	0.238	−0.522**	0.243	−0.532**	0.234
Treatment × Terai and Madeshi	0.087*	0.052	0.105**	0.052	0.066	0.053
Treatment × Dalit	0.050	0.038	0.053	0.038	0.041	0.040
Treatment × Newar	0.146*	0.079	0.148*	0.079	0.123**	0.082
Treatment × Janajati	0.008	0.038	0.015	0.038	−0.012	0.043
Treatment × Muslim	−0.044	0.184	−0.027	0.173	−0.087	0.179
Treatment × other castes	0.040	0.058	0.069	0.056	0.019	0.061
Treatment × time to facility 1 > 4 h	−0.068*	0.041	−0.073*	0.041	−0.081	0.041
Treatment × time to facility 4 > 24 h	−0.029	0.049	−0.025	0.050	−0.047	0.050
Treatment × time to facility > 1 day	−0.048	0.053	−0.056	0.052	−0.071	0.056
Treatment × urban	−0.035	0.060	−0.025	0.059	−0.056	0.060
Treatment × previous delivery SDIP	0.008	0.039	0.005	0.040	0.006	0.040
Treatment × active FCHV	0.067	0.102	0.054	0.104	0.070	0.102
Treatment × active women's groups	0.011	0.028	0.006	0.028	0.012	0.029
Treatment × hill	−0.103*	0.056	–	–	–	–
Treatment × mountain	−0.007	0.051	–	–	–	–
Treatment × ratio of subsidy to cost	–	–	0.205**	0.096	0.996***	0.310
Treatment × hospital quality	–	–	–	–	0.137**	0.059
Treatment × PHCC quality	–	–	–	–	0.085**	0.033
Treatment × health post quality	–	–	–	–	−0.050	0.031
Treatment × year	0.008	0.018	0.011	0.018	0.011	0.018
R^2		0.213		0.211		0.217
F-statistic for null that all interactions = 0		1.81		1.69		2.21
p-Value		0.0072		0.0171		0.0002
F-statistic for null that regional interactions = 0		4.32		–		–
p-value		0.0134		–		–
F-statistic for null that accessibility interactions = 0		1.21		1.51		1.50
p-Value		0.3046		0.209		0.212
F-statistic for null that caste interactions = 0		1.18		1.44		0.93
p-Value		0.3157		0.1936		0.4691

Note: Sample includes 5901 deliveries. Results are based on a linear probability model, which is weighted by the kernel weight provided by `psmatch2`. Each model includes all the covariates that were interacted with the treatment indicator, but the effects of these variables are not reported. FCHV = female community health volunteer. PHCC = primary health care centre.

* Impact parameters are significant at 10%.
** Impact parameters are significant at 5%.
*** Impact parameters are significant at 1%.

and the quality of care is low (Lagarde et al., 2007). We find that the effect of the SDIP on utilisation was positively associated with our proxy for quality of care at hospitals and primary health care centres, which provides support to the argument that demand-side incentives are most effective once supply-side improvements have been put place. In the literature on conditional cash transfers, the issue of interaction between supply and demand has been given almost no attention and is a pressing area for further research, particularly in view of the fact that financial incentives are becoming more widespread in low-income countries. The analysis of heterogeneity also finds that larger financial incentives are associated with a greater impact on utilisation, as expected.

5.1. Interpretation of magnitudes

The treatment group was defined in such a way that the measure of programme impact essentially filters out problems in the implementation. The analysis of the SDIP's promise of cash provides an upper bound on the treatment effect and can be interpreted as the impact on utilisation if there were no problems in the disbursement of cash to beneficiaries. This is useful in separating the question of whether the incentive mechanism underpinning the programme worked from questions about administrative failings. An alternative study design, such as randomisation of the programme across intervention and control districts, would have struggled to detect any impact because implementation was poor. The unusual study design does mean, however, the findings should be accompanied by information on implementation, particularly if the interpretation is to be meaningful to policymakers.[13]

To further interpret the magnitudes of the effect of the SDIP we take into account the extent of implementation. Because coverage of the SDIP was low, the increase in utilisation over the entire sample is a fraction of the estimates given by the ATT. With just

[13] Particular care should be taken in the interpretation of results for subgroups. It is perfectly possible, and indeed likely, that the SDIP had a large impact on utilisation of a subgroup (i.e. high ATT estimate), yet failed to reach many in that subgroup (i.e. low uptake).

over one-quarter of the sample reached by the programme, the ATT estimate implies that the increase in skilled birth attendance as a result of the SDIP was 1.1 percentage points across the entire sample. Similarly, the rise in the utilisation of institutional deliveries attributable to the SDIP is 1.0 percentage points and, in the case of the caesarean section rate, 0.3 percentage points. These effects are modest, particularly when we consider the Government of Nepal's target for skilled birth attendance is 60 percent in 2015.[14]

There is a growing literature on demand-side incentives for health against which to compare the magnitudes of our estimated effects, although few are specific to maternal health. Experimental evidence comes from studies of conditional cash transfers in Mexico (Fernald et al., 2008; Gertler, 2000, 2004), Nicaragua (Maluccio and Flores, 2005), Brazil (Morris et al., 2004b), Ecuador (Paxson and Schady, 2008) and Honduras (Morris et al., 2004a), one-off financial incentives in Malawi (Thornton, 2008), and non-financial incentives in India (Banerjee et al., 2010). The interventions in these studies are targeted towards poor families and most provide some evidence of positive effects on utilisation of health services and immunization coverage.[15]

In Malawi, a small-scale project was found to increase the percentage of individuals who collected their HIV test results by 44 percentage points (Thornton, 2008). Perhaps the most well-known CCT programme is Mexico's *Oportunidades*, which was shown to increase health clinic consultations by 2.1 visits per day (Gertler, 2000). The CCT programme in Honduras increased utilisation of prenatal care by women, routine paediatric examinations and child growth monitoring by 19 percentage points, 20 percentage points and 16 percentage points respectively (Morris et al., 2004a). A similar programme in Nicaragua increased utilisation of child preventive health visits by 11 percentage points (Maluccio and Flores, 2005). No significant impact of CCTs on health visits was found in Ecuador (Paxson and Schady, 2008). Finally, in the Indian state of Rajasthan, lentils were offered alongside immunization camps, raising full immunization rates to 39 percent compared with 6 percent in control and 18 percent in immunization camp only villages (Banerjee et al., 2010). The evidence on financial incentives in health is limited almost exclusively to the use of simple health technologies. More complex health services, whose quality of care is more difficult for patients to assess, have rarely been targeted using financial incentives.

While it is beyond the scope of this paper to conduct a detailed cost-effectiveness analysis,[16] we are able to make a back-of-the-envelope calculation of the cost per additional facility birth.[17] Using information on programme spending, we estimate that the SDIP's expenditure on demand-side incentives amounts to US$ 115 for each additional delivery in a health facility. When we factor in the cost of delivery care using data from Borghi et al. (2006a), we estimate the cost to be approximately US$ 210 per additional facility birth. This should be considered a lower bound estimate since we include no programme administration costs and we almost certainly have not captured the full economic cost of delivery care. Given that the cash payment ranges from $8 in the tarai districts to around $40 in the mountain districts, the estimate of US$ 115 gives some sense of the inefficiency of providing universal financial incentives.

The combination of the two sets of findings on impact and implementation suggest a missed opportunity for the government. Because implementation was far from complete, the financial incentives were unable to bite and in practice few women were incentivised to seek formal care at childbirth.[18] These results imply that the impact of the SDIP on utilisation over the entire population has been too low to generate concerns of maternity services being overwhelmed from an influx of deliveries. Indeed, there was no evidence from the qualitative studies in the evaluation to suggest otherwise. The magnitude of the impact estimates also suggest there can be little expectation of a sizeable improvement

Table A1
Probit model results for the estimation of the propensity score.

Variable	Coefficient	Std error	*z*-Stat
Constant	7.029	5.420	1.30
Age of woman	−0.007	0.004	−1.94
Log of wealth asset score	−14.183	7.565	−1.87
Log of wealth asset score^2	7.596	3.862	1.97
Log of wealth asset score^3	−1.698	0.855	−1.99
Log of wealth asset score^4	0.138	0.069	1.99
Primary education	0.255	0.065	3.94
Secondary education	0.294	0.061	4.78
Higher education	0.529	0.083	6.40
Woman works in agriculture	0.134	0.062	2.18
Woman is salaried or government worker	0.265	0.145	1.82
Woman has small business	0.016	0.099	0.16
Woman is a waged worker	0.102	0.095	1.08
Woman has other work	−0.060	0.374	−0.16
Terai and Madeshi Other castes	0.042	0.083	0.51
Dalit	−0.054	0.064	−0.84
Newar	−0.020	0.124	−0.16
Janajati	−0.198	0.061	−3.25
Muslim	−0.639	0.206	−3.10
Other castes	−0.142	0.081	−1.75
Walk to facility 1 h < 4 h	−0.158	0.057	−2.76
Walk to facility 4 h < 1 day	−0.422	0.076	−5.52
Walk to facility > 1 day	−0.670	0.105	−6.36
Urban	0.125	0.078	1.59
Woman had a previous delivery during SDIP	0.130	0.074	1.76
Active female community health volunteer	2.829	0.192	14.74
Active women's groups	0.161	0.043	3.74
Sankhuwasabha	0.854	0.086	9.99
Myagdi	0.600	0.083	7.25
Rupandehi	0.546	0.098	5.57
Jumla	0.020	0.074	0.27
Achham	0.651	0.104	6.25
Year	0.336	0.023	14.54
N	5905		
Pseudo R^2	0.1765		
Log likelihood	−2745.6		

[14] According to our data, coverage of skilled birth attendance grew from 19 percent to 24 percent over the period 2005–2008, still well below the 60 percent government target.

[15] We report below effects that were shown to be significant but note that the studies also show evidence of no effect on numerous other utilisation outcomes, which are summarized in a systematic review of CCT programmes by Lagarde et al. (2007). Robust estimates from nonexperimental studies include CCT programmes in Columbia (Attanasio et al., 2005), Turkey (Ahmed et al., 2007), and Chile (Galasso, 2007).

[16] In particular, we consider it too speculative to attempt to impute an estimate of lives saved from our utilisation impact results given that there is little or no rigorous evidence on the effect of facility births on health outcomes (Campbell and Graham, 2006).

[17] The cost is based on SDIP expenditure on the demand-side incentives over the period 2005–2008 and estimates of the cost of giving birth in a health facility from Borghi et al. (2006a), adjusted for inflation. We use an exchange rate of 70 Nepalese Rupees per dollar. The cost does not include the cost of SDIP administration, nor the cost of technical assistance provided to the programme by the Support to Safe Motherhood Programme. The effect of the SDIP is calculated on the basis of the 1.0 percentage point increase in the number of institutional deliveries attributable to the demand-side incentives. The total number of deliveries is calculated using population projections and estimates of the crude birth rate from the Census 2001 and the Demographic and Health Survey 2006 respectively.

[18] We can speculate further that public's experience of the programme may have eroded the effectiveness of future policy by undermining trust in the government to meet its stated commitments.

Table A2
Balancing of the covariates by type of matching estimator.

Variable	Before matching		After kernel matching			After mahalanobis matching			After nearest neighbour matching		
	Bias (%)	*t*-Stat	Bias (%)	*t*-Stat	% Reduction in bias	Bias (%)	*t*-Stat	% Reduction in bias	Bias (%)	*t*-Stat	% Reduction in bias
Age of woman	−22.0	−7.13	−1.2	−0.33	94.6	1.6	0.47	92.8	−1.7	−0.48	92.3
Ln(wealth asset score)	22.7	7.52	2.3	0.64	89.8	3.2	0.90	85.7	3.0	0.83	86.9
Ln(wealth asset score)^2	21.2	7.08	2.4	0.66	88.6	3.5	0.96	83.4	3.2	0.87	84.9
Ln(wealth asset score)^3	19.7	6.67	2.4	0.65	87.6	3.7	0.98	81.5	3.3	0.89	83.1
Ln(wealth asset score)^4	18.5	6.31	2.4	0.63	86.9	3.7	0.98	79.9	3.4	0.89	81.6
Primary education	4.9	1.66	−1.2	−0.32	75.2	3.8	1.04	21.7	−2.6	−0.67	47.5
Secondary education	18.1	6.22	−1.7	−0.44	90.6	−1.4	−0.37	92.1	−1.3	−0.33	93.0
Higher education	32.2	11.83	4.9	1.17	84.8	3.8	0.90	88.2	4.3	1.03	86.7
Woman works in agriculture	−3.8	−1.26	−0.3	−0.09	90.7	−3.6	−0.99	3.2	−0.5	−0.12	87.9
Woman is salaried worker	12.3	4.61	3.1	0.75	74.5	0.0	0.00	100.0	1.8	0.42	85.5
Woman has small business	8.6	3.01	−3.5	−0.85	59.7	1.2	0.30	86.5	−4.6	−1.10	47.3
Woman is a waged worker	−8.1	−2.62	−2.8	−0.79	65.6	1.8	0.54	77.7	−0.4	−0.12	94.8
Woman has other work	2.5	0.91	1.0	0.25	61.0	0.0	0.00	100.0	1.0	0.26	58.9
Terai and Madeshi other castes	−6.3	−2.04	−1.3	−0.38	78.9	1.0	0.30	83.4	−2.9	−0.81	53.5
Dalit	−6.9	−2.26	0.6	0.16	92.0	1.6	0.46	76.8	1.7	0.48	75.7
Newar	16.1	6.10	1.0	0.22	94.1	0.0	0.00	100.0	0.2	0.05	98.7
Janajati	−7.9	−2.59	0.2	0.06	97.3	−0.7	−0.19	91.3	0.5	0.14	93.7
Muslim	−18.4	−5.28	−0.2	−0.11	98.8	0.0	0.00	100.0	0.1	0.03	99.7
Other castes	−18.0	−5.69	−1.1	−0.33	94.1	0.2	0.07	98.8	−0.9	−0.28	94.9
Distance to facility 1 > 4 h	8.4	2.81	−0.2	−0.04	98.2	−6.7	−1.86	20.7	0.7	0.20	91.2
Distance to facility 4 > 24 h	−19.6	−6.24	−0.3	−0.10	98.3	2.3	0.70	88.4	−0.8	−0.24	96.0
Distance to facility > 1 day	−29.9	−8.89	−0.3	−0.12	98.9	0.5	0.21	98.2	−0.8	−0.32	97.2
Urban	12.1	4.22	3.2	0.83	73.6	5.1	1.33	57.8	3.6	0.93	70.2
Previous delivery during SDIP	11.5	4.04	1.5	0.37	87.2	11.7	3.20	−1.6	−1.0	−0.26	91.1
Active FCHV	62.1	22.20	−2.9	−0.69	95.3	12.8	3.18	79.3	−4.3	−1.01	93.2
Active women's groups	13.8	4.66	3.6	0.96	73.9	9.5	2.57	31.0	2.7	0.70	80.7
Sankhuwasabha	25.1	8.83	−2.0	−0.51	91.8	1.0	0.26	95.9	−1.8	−0.44	92.9
Myagdi	5.8	1.95	1.3	0.36	76.6	1.3	0.35	77.3	1.0	0.26	83.0
Rupandehi	14.7	5.05	−1.9	−0.48	87.3	0.3	0.09	97.7	−2.6	−0.65	82.6
Jumla	−21.5	−6.82	0.3	0.08	98.8	−1.1	−0.35	94.7	0.8	0.25	96.2
Achham	−16.5	−5.29	3.0	0.92	81.5	−0.4	−0.11	97.8	2.8	0.85	82.9
Year	40.5	13.06	−1.4	−0.38	96.6	22.8	6.64	43.7	−0.1	−0.02	99.9

in maternal health, particularly when one accepts that increased utilisation does not automatically translate into improved health.

5.2. Limitations

Discussion of a number of issues allows consideration of the limitations of the analysis. First, the findings rest on the plausibility of the conditional independence assumption, which is ultimately untestable. We assessed the robustness of the findings to this assumption and on the basis of these results believe we are able to make causal inferences from our findings. In further support of this conclusion, we would argue that internal validity is strengthened by the richness of the data (Heckman et al., 1998; Diaz and Handa, 2006) and the fact that selection bias may be less of a problem in the context of this study because individuals do not actively select or enrol into the programme. Rather, selection into treatment is a passive process and it seems reasonable to assume that selection bias on unobservables is more of a problem in programmes which require a decision on the part of the beneficiary to participate, since the decision may be influenced by, rather than simply correlated with, the unobservables.

Second, the sample was not designed to be representative of the entire country, which means the results should not be interpreted as national impact estimates. Third, our study covered only the early period of the SDIP and all indications are that the programme has improved since our household survey. Finally, our analysis was limited to utilisation outcomes as the main measure of performance. In contrast with many simple health technologies, the link between utilisation of maternity services and improved health depends on a complex set of factors linked to the quality of care. Individuals demand health, not health care per se. Judging the success of the SDIP on the extent to which it influenced health seeking behaviour at childbirth is thus fraught with problems. The standard measure of utilisation, skilled birth attendance, is a crude measure of access to quality maternity services. Not only does it fail to capture well the skills of the attendant, it says nothing about many other important aspects of quality that impact on health. Ideally, we would have measured maternal deaths as the main outcome of interest but because they are such a rare event the sample size would have been prohibitively large to detect even a large effect.[19]

6. Conclusion

In this paper we find that a national programme offering financial incentives to households was modestly effective in encouraging women to deliver with professional care in six districts of Nepal. The impacts appear to be modified by the size of the financial package relative to the cost of care and the quality of care provided in hospitals and primary health care centres. Owing to the low coverage of the SDIP, a small proportion of the population were incentivised by the programme to seek formal care at childbirth, illustrating how slow implementation has constrained its success.

[19] This point is well illustrated by a randomised experiment of vitamin A supplementation on maternal survival in Ghana which lasted seven years and required over 200,000 participants to conclude that the intervention had no effect (Kirkwood et al., 2010).

Table A3
Impact estimates by matching procedure.

Variable	Kernel		Mahalanobis		Nearest neighbour	
	ATT	*t*-Stat	ATT	*t*-Stat	ATT	*t*-Stat
Place of delivery						
Health facility	0.040***	2.70	0.039**	2.06	0.046***	3.04
Government health facility	0.043***	3.28	0.045**	2.65	0.048***	3.52
NGO hospital	−0.011*	−1.86	−0.009	−1.17	−0.011*	−1.73
Private health facility	0.008	1.40	0.003	0.48	0.009	1.57
Type of attendant						
Doctor, nurse or midwife (SBA)	0.042***	2.72	0.042**	2.11	0.048***	3.00
Any professional health worker	0.052***	3.17	0.045**	2.16	0.058***	3.43
Procedure at delivery						
Caesarean section	0.012*	1.83	0.018**	2.13	0.013*	1.83
Caesarean section or assisted	0.019**	2.02	0.032***	2.67	0.022**	2.25

Note: NGO = non-governmental organisation; SBA = skilled birth attendant.
* Significance at 10%.
** Significance at 5%.
*** Significance at 1%.

Acknowledgements

We have benefited enormously from discussions with Anthony Costello, Anne Mills, Suresh Tiwari, Basu Dev Neupane, Louise Hulton, Greg Whiteside and seminar participants at iHEA, Oxford University and University College London. We thank the dedicated group of individuals involved in the data collection. The study was funded by the Department for International Development through the Support to Safe Motherhood Programme in Nepal (managed by Options UK). Financial support from the ESRC and MRC is gratefully acknowledged. The research was approved by the ethics committee of the London School of Hygiene and Tropical Medicine. All errors are our own.

Appendix A.

See Tables A1–A3.

References

Ahmed, A., Adato, M., Kudat, A., Gilligan, D., Roopnaraine, T., Colasan, R., 2007. Impact Evaluation of the Conditional Cash Transfer Program in Turkey: Final Report. International Food Policy Research Institute, Washington, DC.
Angrist, A.A., Pischke, J., 2009. Mostly Harmless Econometrics: An Empiricist's Companions. Princeton University Press, Princeton.
Attanasio, O., Gómez, L., Heredia, P., Vera-Hernández, M., 2005. The Short-term Impact of a Conditional Cash Subsidy on Child Health and Nutrition in Colombia. Institute of Fiscal Studies, London.
Banerjee, A.V., Duflo, E., Glennerster, R., Kothari, D., 2010. Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. BMJ 340, c2220.
Borghi, J., Ensor, T., Neupane, B.D., Tiwari, S., 2006a. Financial implications of skilled attendance at delivery in Nepal. Tropical Medicine and International Health 11, 228–237.
Borghi, J., Ensor, T., Somanathan, A., Lissner, C., Mills, A., 2006b. Mobilising financial resources for maternal health. Lancet 368, 1457–1465.
Campbell, O.M., Graham, W.J., 2006. Strategies for reducing maternal mortality: getting on with what works. Lancet 368, 1284–1299.
Cochran, W.G., Rubin, D.B., 1973. Controlling bias in observational studies. Sankhya – The Indian Journal of Statistics Series A 35, 417–446.
Diaz, J.J., Handa, S., 2006. An assessment of propensity score matching as a non-experimental impact estimator—Evidence from Mexico's PROGRESA program. Journal of Human Resources 41, 319–345.
Djebbari, H., Smith, J., 2008. Heterogeneous impacts in PROGRESA. Journal of Econometrics 145, 64–80.
Ensor, T., Clapham, S., Prasai, D.P., 2009. What drives health policy formulation: insights from the Nepal maternity incentive scheme? Health Policy 90, 247–253.
Fernald, L.C., Gertler, P.J., Neufeld, L.M., 2008. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. Lancet 371, 828–837.
Filmer, D., Pritchett, L., 1998. Estimating Wealth Effects with No Expenditure Data – Or Tears: An Application to Educational Enrollment in States of India. World Bank, Washington, DC.
Fiszbein, A., Schady, N., Ferreira, F.H., Grosh, M., Kelleher, N., Olinto, P., Skofias, E., 2008. Conditional Cash Transfers: Reducing Present and Future Povertys. IBRD/The World Bank.
Galasso, E., 2007. With their Effort and One Opportunity: Alleviating Extreme Poverty in Chile. World Bank, Washington, DC.
Gertler, P., 2000. Final Report: The Impact of PROGESA on Health. International Food Policy Research Institute, Washington, DC.
Gertler, P., 2004. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. American Economic Review 94, 336–341.
Gertler, P., Locay, L., Sanderson, W., 1987. Are user fees regressive? The welfare implications of charging for medical care in Peru. Journal of Econometrics 36 (supp), 67–68.
Gertler, P., Van der Gaag, J., 1990. The Willingness to Pay for Medical Care: Evidence from Two Developing Countries. Johns Hopkins University Press, Baltimore and London.
Government of Nepal, 2001. Nepal Demographic and Health Surveys. ORC Macro, MD, USA/Kathmandu, Nepal.
Government of Nepal, 2005. Operational guidelines on incentives for safe delivery services. Government of Nepal, Ministry of Health and Population, Kathmandu.
Government of Nepal, 2007. Nepal Demographic and Health Survey. ORC Macro, MD, USA/Kathmandu, Nepal.
Heckman, J., Ichimura, H., Todd, P., 1998. Matching as an econometric evaluation estimator. Review of Economic Studies, 261–294.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–475.
Imbens, G.W., Wooldridge, J.M., 2009. Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature 47, 5–86.
Jack, A., 2008. World Bank backs anti-Aids experiment. Financial Times, London.
Kirkwood, B.R., Hurt, L., Amenga-Etego, S., Tawiah, C., Zandoh, C., Danso, S., Hurt, C., Edmond, K., Hill, Z., Ten Asbroek, G., Fenty, J., Owusu-Agyei, S., Campbell, O., Arthur, P., 2010. Effect of vitamin A supplementation in women of reproductive age on maternal survival in Ghana (ObaapaVitA): a cluster-randomised, placebo-controlled trial. Lancet 375, 1640–1649.
Koblinsky, M., Matthews, Z., Hussein, J., Mavalankar, D., Mridha, M.K., Anwar, I., Achadi, E., Adjei, S., Padmanabhan, P., Marchal, B., De Brouwere, V., van Lerberghe, W., 2006. Going to scale with professional skilled care. Lancet 368, 1377–1386.
Lagarde, M., Haines, A., Palmer, N., 2007. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. JAMA 298, 1900–1910.
Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testings.
Maluccio, J., Flores, R., 2005. Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social. International Food Policy Research Insitute, Washington, DC.
Morris, S., Flores, R., Olinto, P., Medina, J.M., 2004a. Monetary incentives in primary health care and effects on use and coverage of preventive health care interventions in rural Honduras: cluster randomised trial. Lancet 364, 2030–2037.
Morris, S.S., Olinto, P., Flores, R., Nilson, E.A., Figueiro, A.C., 2004b. Conditional cash transfers are associated with a small reduction in the rate of weight gain of preschool children in northeast Brazil. Journal of Nutrition 134, 2336–2341.
NORA.D, 2007. The Global Campaign for the Health Millennium Development Goalss. NORAD, Oslo.
Oxman, A., Fretheim, A., 2008. An Overview of Research on the Effects of Results-based Financing vol. Report Nr 16-2008. Nasjonalt kunnskapssenter for helsetjenest, Oslo.

Paxson, C., Schady, N., 2008. Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador. World Bank, Washington, DC.
Powell-Jackson, T., Morrison, J., Tiwari, S., Neupane, B.D., Costello, A.M., 2009a. The experiences of districts in implementing a national incentive programme to promote safe delivery in Nepal. BMC Health Servies Research 9, 97.
Powell-Jackson, T., Neupane, B., Tiwari, S., Tumbahangphe, K., Manandhar, D., Costello, A., 2009b. The impact of Nepal's national incentive progamme to promote safe delivery in the district of Makwanpur. In: Chernichovsky, D., Hanson, K. (Eds.), Innovations in Health System Finance in Developing and Transitional Economies. Emerald, Bingley.
Pradhan, A., Aryal, R., Regmi, G., Ban, B., Govindasamy, P., 1997. Nepal Family Health Survey 1996. Government of Nepal, Ministry of Health/Macro International Inc., New Era/Calverton, Maryland and Kathmandu.
Rohde, J., Cousens, S., Chopra, M., Tangcharoensathien, V., Black, R., Bhutta, Z.A., Lawn, J.E., 2008. 30 years after Alma-Ata: has primary health care worked in countries? Lancet 372, 950–961.
Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41.
Rosenbaum, P., Rubin, D., 1984. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 516–524.
Rosenbaum, P., Rubin, D., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician 39, 33–38.
Sargan, J., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26, 393–415.
Smith, J.A., Todd, P.E., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125, 305–353.
Thornton, R.L., 2008. The demand for, and impact of, learning HIV status. American Economic Review 98, 1829–1863.
UNICEF, 1997. Guidelines for Monitoring the Availability and Use of Obstetric Servicess. UNICEF, New York.
Wagstaff, A., Lindelow, M., Jun, G., Ling, X., Juncheng, Q., 2009. Extending health insurance to the rural population: an impact evaluation of China's new cooperative medical scheme. Journal of Health Economics 28, 1–19.
Weiss, C., 1998. Evaluation: Methods for studying programs and policies. Prentice Hall.
White, H., Masset, E., 2007. Assessing interventions to improve child nutrition: a theory-based impact evaluation of the Bangladesh Integrated Nutrition Project. Journal of International Development 19, 627.