Should Aid Reward Performance?

Evidence from a field experiment on health and education in Indonesia

Benjamin A. Olken, MIT
Junko Onishi, World Bank
Susan Wong, World Bank

August 2011

ABSTRACT

This paper reports an experiment in over 3,000 Indonesian villages designed to test the role of performance incentives in improving the efficacy of aid programs. Villages in a randomly-chosen one-third of subdistricts received a block grant to improve 12 maternal and child health and education indicators, with the size of the subsequent year's block grant depending on performance relative to other villages in the subdistrict. Villages in remaining subdistricts were randomly assigned to either an otherwise identical block grant program with no financial link to performance, or to a pure control group. We find that the incentivized villages performed better on health than the non-incentivized villages, particularly in less developed provinces, but found no impact of incentives on education. We find no evidence of negative spillovers from the incentives on untargeted outcomes. Incentives led to what appear to be more efficient use of block grants, and led to an increase in labor from health providers, who are partially paid fee-for-service, but not teachers. On net, between 50-75% of the total impact of the block grant program on health indicators can be attributed to the performance incentives.

We thank the members of the PNPM Generasi Team including: Sadwanto Purnomo, Gerda Gulo, Juliana Wilson, Gregorius Pattinasarany, Scott Guggenheim, John Victor Bottini, and Sentot Surya Satria. Special thanks go to Yulia Herawati, Gregorius Endarso, Joey Neggers, and Lina Marliani for their outstanding support in survey preparation, oversight and research assistance. We thank the Government of Indonesia through the Ministry of Planning (Bappenas), the Coordinating Ministry for Economy and Social Welfare (Menkokesra), and the Ministry of Home Affairs (Depdagri) for their support for the program and its evaluations. Special thanks to Sujana Royat (Menkokesra); Prasetijono Widjojo, Endah Murniningtyas, Pungky Sumadi, Vivi Yulaswati (Bappenas); and Ayip Muflich, Eko Sri Haryanto, and Bito Wikantosa (Ministry of Home Affairs). The University of Gadjah Mada (UGM), Center for Public Policy Studies, implemented the surveys used in this analysis. Financial support for the overall PNPM Generasi program and the evaluation surveys has come from the Government of Indonesia, the World Bank, the Decentralization Support Facility, the Netherlands Embassy, and the PNPM Support Facility, which consists of donors from Australia, the United Kingdom, the Netherlands, and Denmark, and the Spanish Impact Evaluation Fund. The views expressed in this paper are those of the authors alone and do not represent the views of the World Bank or any of the many individuals or organizations acknowledged here.

1. Introduction

A recent movement throughout the world has sought to improve the links between development aid and performance. For example, the United Nations has sought to focus developing country governments on improving human development and poverty alleviation by defining and measuring progress against the Millennium Development Goals. Even more directly, foreign assistance given out by the U.S. Millennium Challenge Corporation is explicitly conditioned on recipient countries meeting 17 indicators of good governance, ranging from civil liberties to immunization rates to girl's primary education rates to inflation (Birdsall and Savedoff 2009). The World Bank is similarly moving towards "Program for Results" loans, which would condition actual World Bank disbursements on results obtained. The idea of linking aid to performance is not limited to the developing world: the U.S. has used a similar approach to encourage state-based education and local school performance reform through its Race To The Top and No Child Left Behind programs.

As in any principal-agent framework, linking aid to performance can be useful from the principal's perspective to the extent it creates incentives for lower tiers of government to improve effort and mobilize additional resources. But there are potential pitfalls as well. As with all incentive schemes, there can be multitasking problems, where effort allocated towards targeted indicators comes at the expense of other, non-incentivized indicators that the principal may also care about (Holmstrom and Milgrom 1991). There can also be direct attempts to manipulate indicators to increase payouts (Linden and Shastry 2008). And – particularly when government budgets are being allocated based on performance– there is a substantial risk that performance-based aid will mean that budgets are directed to richer or otherwise better performing locations. If these richer or better performing regions have a lower marginal value of funds, this reallocation offset could offset some of the incentive effects.

This paper investigates these issues in the context of a new aid program in Indonesia that provides block grants to villages to improve health and education and incorporates explicit performance incentives. We designed a randomized, field experiment to test the impact of pay-for-performance as part of the Indonesian National Community Empowerment Program – Healthy and Smart Generation (*Program National Pemberdayaan Masyarakat-Generasi Sehat dan Cerdas*; henceforth referred to as *Generasi*), which began in 2007. Under the program, villages received an annual block grant, which each village could allocate to any activity that supported one of 12 indicators of health and education service delivery (such as prenatal and postnatal care, childbirth assisted by trained personnel, immunizations, school enrollment, and school attendance). To give communities incentives to focus on the most effective policies, 20% of the subsequent year's block grant is allocated among villages in a subdistrict based on their relative performance on each of the 12 targeted health and education indicators. The *Generasi* program was designed to take the idea of performance incentives for health and education from Conditional Cash Transfer (CCTs) programs, such as Mexico's Progresa program (Gertler 2004; Schultz 2004), and apply it at the village level to allow communities the flexibility to address supply constraints, demand constraints, or some combination.

Unlike most evaluations of CCTs, which cannot separately identify the impact of the incentives from the impact of the additional cash provided, the *Generasi* evaluation was designed to specifically isolate the impact of the incentives from the overall impact of the program.[1] Specifically, the Government of Indonesia, working in consultation with the authors, designed two versions of the Generasi program, the performance based program designed above and an otherwise completely identical version of the program without financial incentives. A total of

[1] Notable exceptions are two recent studies that seek to unpack the incentive effects from traditional CCTs using a similar type of design. See Baird, McIntosh, and Ozler (2011), and Devoto, Duflo, and Dupas (in progress).

264 subdistricts, each consisting of approximately 12 villages, were randomized into either a pure control group or one of two versions of the program: the “incentivized” version with the pay-for-performance component, or the otherwise identical, “non-incentivized” version without the pay-for-performance incentives. Evaluation surveys were conducted at baseline, 18 months after the program started, and 30 months after the program started. With over 2,100 villages randomized to receive the *Generasi* program (plus over 1,000 villages in control subdistricts), and over 1.8 million target beneficiaries of the program in treatment areas, to the best of our knowledge this represents one of the largest randomized social experiments conducted in the world to date.

We begin by examining the impact of the incentives on the 12 main indicators the program was designed to address. Using data from the household survey, we find that the incentives led to greater performance on the health indicators, but not on the education indicators. Over the two years of the program, the 8 targeted maternal and child indicators (e.g., prenatal visits, delivery by trained midwives, childhood immunizations, growth monitoring) were an average of 0.03 standard deviations higher in incentivized areas than in non-incentivized areas. This was driven by prenatal visits, which were 5% higher in incentivized areas than non-incentivized areas, and immunization rates, which were 3% higher in incentivized areas than non-incentivized areas. While these differences are modest, the impact of the incentives was more pronounced in areas with low baseline levels of service delivery: the incentives improved the 8 targeted maternal and child health indicators by an average of 0.06 standard deviations for a subdistrict at the 10^{th} percentile at baseline, and by as much as 0.10 – 0.15 standard deviations in the poorer, off Java provinces. On net, between 50-75% of the program’s impact on health indicators came from the incentivized areas: for example, on average the 8 target health indicators were 0.052 standard deviations higher than pure controls in incentivized areas, compared to only 0.019 standard deviations higher than controls in non-incentivized areas. There were no differences between incentivized and non-incentivized areas on the 4 education indicators examined (primary and junior secondary enrollment and attendance).

We find relatively little evidence of adverse effects of the incentives. We find no evidence of a multi-tasking problem across the wide variety of non-incentivized metrics we examine: incentive areas were comparable to non-incentivized areas in terms of quality of health care services, use of adult health care services, good childcare and parenting practices, high school enrollment, enrollment in alternative forms of education, and child labor. We find no evidence that immunization recordkeeping or school attendance was manipulated in performance zones relative to non-performance incentive zones.

We investigate four potential mechanisms through which the incentives may have had an impact: changes in the composition of spending, worker effort, community effort, and targeting of benefits within the community. We find two main channels through which the incentives may have had an impact. First, the incentives led to a reallocation of the block grants away from school supplies and uniforms (4 percentage points, or about 16 percent) and towards health expenditures (3 percentage points, or about 6.5 percent). Despite the reallocation of funds away from school supplies and uniforms, households were in fact no less likely to receive these items, and were more likely to receive education scholarships in the performance areas. This suggests that the changes in budgets may have been towards less expensive uniforms and supplies rather than cutting down on quantity. Second, we find that the performance incentives led to an increase in the labor of midwives, who are the major providers of the incentivized maternal and child health services (1.7 hours over 3 days, or about 6 percent). By contrast, we found no change in labor supplied by teachers. One possible explanation is that midwives are paid on a fee for service basis for many of the services they provide (e.g., pre and post natal care, deliveries), whereas teachers are not.

We perform a standard cost-effectiveness calculation to help interpret the magnitudes of the impacts we find. We use the weights assigned by the program to each of the 12 indicators for calculating bonus points to calculate an average dollar cost per "bonus point" achieved. On average, we find that the Generasi program cost about $8 - $11 to generate one additional "bonus point." Translating "bonus points" back into outcomes suggests, for example, that the implied cost of additional child weight check was $16 - $22, the cost of preventing one malnourished child was $384 - $528, the cost of getting one additional child fully covered with Vitamin A was $160 - $220, and the cost of enrolling one more child in primary school was $200 - $275. These costs are similar to a traditional conditional cash transfer program implemented in Indonesia at the same time and evaluated using comparable methodologies, though are substantially more expensive than several recent interventions in Kenya and India aimed at improving school enrollment by improving health or providing school inputs.

These cost-effectiveness numbers include the entire cost of the program. If we compare the program with and without performance incentives to isolate the incremental cost of the performance incentives, we obtain numbers on the order of $0.60 per point. This suggests that while the program itself is not particularly cost effective, conditional on doing a block grant program of this type, the incentives should be included, and they suggest that adding similar performance incentives to other, pre-existing block grant schemes (holding the total amount of money spent) could be a very cost effective way to improve aid effectiveness.

This study forms part of a newly expanding literature that seeks to separately identify the impact of performance incentives on health and education, holding the amount of resources given constant. In the context of conditional cash transfers programs, Baird, McIntosh, and Ozler (2011) find that adding conditions to the CCT reduced school dropouts and improved English comprehension. In health, Basinga et al (2011), which conducts a randomized experiment on pay-for-performance for health centers in Rwanda, find positive impacts of performance incentives on institutional deliveries, preventive health visits for young children, and quality of prenatal care, but not on the quantity of prenatal care or immunizations. This study is the first of its kind to study the effect of performance incentives tied to aid to an entire community.

The remainder of the paper is organized as follows. Section 2 discusses the design of the program, the experimental design, and the econometric approach. Section 3 presents the main results of the impact of the incentives on the 12 targeted indicators. Section 4 examines the potential adverse affects of incentives, and Section 5 examines the mechanisms through which the incentives may have acted. Section 6 discusses a cost-effectiveness calculation, and Section 7 concludes.

2. Program and experimental design

2.1. The Generasi Program

This section describes the *Generasi* program, the Indonesian community block grant program that is the focus of this study. To the best of our knowledge, *Generasi* is the first health and education program worldwide that combines community block grants with explicit performance bonuses for communities.

The *Generasi* program began in mid 2007 in rural areas of five Indonesian provinces: West Java, East Java, North Sulawesi, Gorontalo, and Nusa Tenggara Timur. In 2007, the

Generasi program covered 1,605 villages in 129 subdistricts, with a total budget of US$20 million. In the program's second year, which began in mid 2008, the program expanded to cover a total of 2,120 villages in a total of 176 subdistricts, with a total budget of US$44 million. The program is currently ongoing.

The *Generasi* program is oriented around the 12 indicators of maternal and child health behavior and educational behavior shown in column (1) of Table 1. These indicators were chosen by the government of Indonesia to be as similar as possible to the conditions for a conditional cash transfer being piloted at the same time as *Generasi* (but in different locations), and, as such, they are in the same spirit as the conditions used by CCTs in other countries, such as Mexico's *Progresa* (Levy 2006). These 12 indicators represent health and educational behaviors that are within the direct control of villagers—such as, the number of children who receive immunizations, prenatal and postnatal care, and the number of children enrolled and attending school—rather than long-term outcomes, such as test scores or infant mortality. They also correspond to the Indonesian government's standard of service for maternal and child health (Government of Indonesia, 1997, 2004) and to the Indonesian government's stated goal of universal primary and junior secondary education (Government of Indonesia, 2008).

In *Generasi,* each year all participating villages receive a block grant to improve maternal health, child health, and education. Block grants are usable for any purpose that the village can claim might help address one of the 12 indicators shown in Table 1, including, but not limited to, hiring extra midwives for the village, subsidizing the costs of prenatal and postnatal care, providing supplementary feeding, hiring extra teachers, opening a branch school in the village (*kelas jauh* or satellite classrooms, or *sekolah terbuka* or formal part-time schooling), providing scholarships, providing school uniforms, providing transportation funds for health care or school attendance, improving health or school buildings, or even building a road or path through the forest to improve access to health and educational facilities. The block grants averaged US$8,500 in the first year of the program and US$13,500 in the second year of the program, or about US$2.70 – US$4.30 per person living in Generasi villages in the target age ranges.

To decide on the allocation of the funds, trained facilitators help each village elect an 11-member village management team, as well as select local facilitators and volunteers. Through social mapping and in-depth discussion groups, villagers identify problems and bottlenecks in reaching the 12 indicators. Inter-village meetings and consultation workshops with local health and education service providers allow community leaders to obtain information, technical assistance, and support from the local health and education offices as well as to coordinate the use of *Generasi* funds for multi-village projects. Following these discussions, the 11-member management team makes the final *Generasi* budget allocation.

A critical (and unique) element of *Generasi* is performance incentives: the size of a village's block grant depends in part on its performance on the 12 targeted indicators. The purpose of the performance bonus is to increase the village's effort at achieving the targeted indicators (Holmstrom 1979), both by encouraging a more effective allocation of *Generasi* funds and by stimulating village outreach efforts to encourage mothers and children to obtain appropriate health care and increase educational enrollment and attendance.

The performance bonus is structured as relative competition between villages within the same subdistrict (*kecamatan*). By making the performance bonuses relative to other villages in the subdistrict, the government sought to minimize the impact of unobserved differences in the capabilities of different areas on the performance bonuses (Lazear and Rosen 1981; Mookherjee 1984; Gibbons and Murphy 1990).

The specific rule for allocating *Generasi* funds is as follows. The size of the overall *Generasi* allocation for the entire subdistrict is predetermined by the subdistrict's population and poverty level.[2] Within a subdistrict, in year 1 of the program funds are divided among villages in proportion to the number of target beneficiaries in each village (i.e., the number of children of varying ages and the expected number of pregnant women). Starting in year 2 of the program, 80 percent of the subdistrict's funds continue to be divided among villages in proportion to the number of target beneficiaries; the remaining 20 percent of the subdistrict's funds form a performance bonus pool, to be divided among villages based on their performance on the 12 *Generasi* indicators.

The performance bonus pool is allocated to villages in proportion to a weighted sum of each village's performance above a predicted minimum achievement level. Specifically, each village's share of the performance bonus pool is determined by:

$$ShareOfBonus_v = P_v / (\Sigma P_j) \text{ where } P_v = \Sigma[\, w_i \times (y_{vi} - m_{vi})]$$

In this formula, y_{vi} represents village v's performance on indicator i, w_i represents the weight for indicator i, m_{vi} represents the predicted minimum achievement level for village v and indicator i, and P_v is the total number of bonus "points" earned by village v.

Generasi uses performance relative to a constant predicted minimum attainment level, rather than improvements over an actual baseline, to avoid the ratchet effect (Weitzman 1980), as well as to avoid the problems inherent in collecting reliable baseline data on performance on all indicators in all villages before the program began. For each of the 12 *Generasi* indicators i, the program set the predicted minimum attainment level, m_{vi}, in village v to be equal to 70 percent of

---

[2] In 2007 the average block grant for each subdistrict was USD 112,300 per subdistrict; in 2008, the average block grant was raised to USD 200,000 per subdistrict. A subdistrict contains roughly between 15,000 and 50,000 individuals and 10 to 20 villages.

the average achievement level for villages with similar levels of access to health and education providers and numbers of beneficiaries, calculated from the 2004 SUSENAS household survey and 2003 PODES census of villages.[3] The minimum level was set at 70 percent of the average to ensure that virtually all villages would be above the minimum threshold, and thus eligible for incentive payments, on each indicator. The weights for each indicator, w_i, were set by the government to be approximately proportional to the marginal cost of having an additional individual complete that indicator. The weights, along with the performance metric for each indicator *i*, are shown in Table 1.

To monitor achievement of the health indicators, facilitators collect data from health providers and community health workers on the amount of each type of service provided. School enrollment and attendance data are obtained from the official school register.[4]

As discussed above, the government implemented two versions of the *Generasi* program to separate the impact of the performance incentives *per se* from the overall impact of having additional financial resources available for health and education: the program with performance bonuses described above (referred to as "incentivized"), and an identical program without performance bonuses (referred to as "non-incentivized"). The non-incentivized version is identical to the incentivized version except that in the non-incentivized version, there is no

---

3 For all health indicators except monthly weighings, access to providers was divided into three categories: 1) having a midwife practicing in the village, 2) not having a midwife in the village but having a midwife practicing within 4km from the center of the village, or 3) not having a midwife practicing within 4km of the village center. For middle school, access was divided into three categories: 1) having a middle school located in the village or within 4km of the village center, 2) having a middle school located between 5 and 9km of the village center, or 3) having a middle school located 10km or more from the village center. For monthly weighings and primary school, all villages were assumed to have the same level of access, since weighing of children is always conducted in the village at monthly *Posyandu* meetings and since all virtually all villages in Indonesia have a primary school.

4 Obtaining attendance data from the official school register is not a perfect measure, since it is possible that teachers could manipulate student attendance records to ensure they cross the 85 percent threshold (Linden and Shastry 2008). While more objective measures of monitoring attendance were considered, such as taking daily photos of students (as in Duflo, Hanna and Ryan 2008) or installing fingerprint readers in all schools (Express India News Service 2008), *Generasi* decided not to adopt these more objective measures due to their cost and logistical complexity. We test for this type of differential manipulation below.

performance bonus pool; instead, in all years, 100 percent of funds are divided among villages in proportion to the number of target beneficiaries in each village. In all other respects, the two versions of the program are identical: the total amount of funds allocated to each subdistrict is the same in both versions, the same socialization materials and indicators are used, the same procedures are used to pick village budget allocations, and the same monitoring tools and scoring system are used. Even the annual points score of villages P_v is also calculated in non-incentivized areas; the only difference is that in non-incentivized villages the points are used simply as an end-of-year monitoring and evaluation tool, and have no relationship to the allocation of funds. Within a given subdistrict, all villages participate in the same version of the program (i.e., either all villages received incentivized *Generasi* or all villages received non-incentivized *Generasi*).

2.2. Experimental design

In order to evaluate the overall impact of *Generasi*, as well as to separately identify the impact of *Generasi*'s performance incentives, *Generasi* locations were selected by lottery to form a randomized, controlled field experiment. The *Generasi* randomization was conducted at the subdistrict (*kecamatan*) level, so that all villages within the subdistrict either received the same version of *Generasi* (either all incentivized or all non-incentivized) or were in the control group. Randomizing at the subdistrict level is important since many health and education services, such as community health centers (*Puskesmas*) and junior secondary schools, provide services to multiple villages within a subdistrict. By randomizing at the subdistrict level, so that all villages in the subdistrict receive the same treatment status, the evaluation design ensures that we capture the total net effect of the program, since any within-subdistrict spillovers would also be captured in other treatment villages..

The *Generasi* locations were selected through the following procedure. First, 300 target subdistricts were identified, targeting poor, rural areas that had an existing community-driven development infrastructure.[5] Locations were spread among five provinces from three different parts of the country: East Java and West Java (these are Indonesia's two most populous provinces, together containing about 35 percent of Indonesia's population), NTT (a relatively poor, remote set of islands in Southeastern Indonesia, typical of Indonesia's small island areas), and Gorontalo and North Sulawesi (located on Sulawesi, one of the three major islands other than Java). On net, the selected 300 subdistricts look broadly similar to all of rural Indonesia on the 12 targeted indicators.[6]

Each of these 300 subdistricts was then randomly assigned by computer into one of three equal-sized groups: incentivized *Generasi* (100 subdistricts), non-incentivized *Generasi* (100 subdistricts), or control (100 subdistricts). Within a subdistrict, all villages received the same treatment. The randomization was stratified by district (*kabupaten*), to ensure a balanced randomization across the 20 different districts in the study. Note that a total of 36 out of the 300 subdistricts in fact should not have been include in the randomization as they were ineligible for Generasi because they had been selected (prior to the randomization) to receive other programs or had had prior implementation problems with previous PNPM programs**.** Since the eligibility

5 Within these five provinces, the government eliminated the wealthiest 20 percent of districts (kabupaten), determined by the district's poverty rate, malnutrition rate, and junior secondary school transition rate, as well as the 28 percent of districts where the PNPM rural infrastructure project was not scheduled to operate in 2007. Since Generasi is implemented through the national PNPM program, it could only be implemented in districts that were already included in the PNPM program. Twenty districts were randomly selected from the remaining eligible districts, stratified by island group. Within the twenty selected districts, subdistricts were eligible for Generasi if they had previously received the PNPM program or were considered less than 67 percent urban by the Central Statistics Office.

6 We compare the 300 Generasi subdistricts to the rest of rural Indonesia, excluding the conflict areas of Aceh and Papua, using the 2004 SUSENAS (i.e., before the Generasi program began). We find that the primary enrollment rate is 96 percent in Generasi areas and 96 percent in non-Generasi areas. The junior secondary enrollment rates are 58 percent in Generasi and 61 percent in non-Generasi areas. For delivery by trained attendant, the rate is 48 percent in both Generasi and non-Generasi areas. The only substantial differences are that the rate of completed immunizations is 29 percent in Generasi areas vs. 22 percent in non-Generasi areas, and the rates for monthly weighings of the youngest child under 5 are 59 percent in Generasi areas vs. 48 percent in non-Generasi areas.

decision was made on the basis of lists determined prior to the randomization, and since we have obtained those lists for treatment and control areas, we exclude ineligible subdistricts in both treatment and control groups from our main analysis.[7]

The Generasi program was phased in over two years. In phasing in the program in the first year (2007), the government for budgetary reasons prioritized those locations that had previously participated in the PNPM rural infrastructure program (denoted group P), since those locations already had the legal infrastructure for distributing PNPM funds and it was easier to re-budget other monies to fund Generasi in those areas. After all group P subdistricts randomized to receive the program had been funded, the government held another lottery to select which group NP subdistricts would being receiving the program in 2007 and which would begin in 2008.[8] By year two of the program (2008) 96% of eligible subdistricts – 174 out of the 181 eligible subdistricts randomized to receive Generasi – were receiving the program. The remaining 7 eligible districts received the regular PNPM program instead of Generasi.[9] The phase-in and final allocation of Generasi is shown in Table 2.

An important consideration for the analysis is the potential for differential provision of other programs in the pure control groups. The main potential avenue through which this might

[7]The determination that these subdistricts would be ineligible had been made prior to the randomization, but was not communicated to the study team, which is why they were included in the randomization. Subdistricts were deemed ineligible if they had been allocated to receive the urban poverty program (UPP), conflict area poverty program (SPADA), or if they had had a previous problem with PNPM implementation. We subsequently obtained the pre-randomization lists used to make this determination, and use these pre-randomization lists to restrict our sample (in both treatment and control areas) to those subdistricts that would actually be eligible for the program. Nevertheless, data collection surveys were conducted in all 300 subdistricts that were initially included in the randomization, regardless of the final eligibility, so as a robustness check we can alternatively estimate intent-to-treat effects using the full 300 subdistricts from the original randomization.

[8] Specifically, in 2007 all 105 eligible group P subdistricts were funded. In group NP, in 2007 *Generasi* was funded in 22 eligible subdistricts. Of these 22 subdistricts, 21 were chosen randomly by computer, stratified by province, in a second lottery among Group NP locations. Group P status was determined prior to randomization. All but 7 of the remaining NP subdistricts were added in 2008.

[9] We do not know why these 7 districts received regular PNPM rather than Generasi. We therefore include them in the treatment group as if they had received the program, and interpret the resulting estimates as intent-to-treat estimates.

occur is other PNPM programs. Specifically, to ensure a fair allocation of funds, the Ministry of Home Affairs decided that no subdistrict would receive both the *Generasi* program and other PNPM programs, which typically fund local infrastructure (roads, bridges, etc.) and microcredit. In 2007, 17 (out of 83) eligible control subdistricts received other PNPM programs, while no treatment subdistricts did; in 2008, 31 (out of 83) eligible control subdistricts received other PNPM programs, as did the 7 eligible subdistricts did that should have been receiving Generasi in 2008 but received regular PNPM rural instead. Since regular PNPM programs tend to focus on basic infrastructure, not health and education, it is unlikely that the differential provision of other PNPM programs in control areas will have substantial impacts on the *Generasi* evaluation results. Nevertheless, in interpreting the results, it is important to recognize that the some portion of the eligible "pure control" subdistricts received PNPM.

2.3. Data

The main data for the analysis is from a set of surveys of households, village officials, health service providers, and school officials. Three waves of the survey were conducted as part of the evaluation series. Wave I, the baseline round, was conducted from June to August 2007 prior to Generasi implementation. Wave II, the first follow-up survey round, was conducted from October to December 2008. Wave III, a longer-term follow-up survey round, was conducted from October 2009 to January 2010. These surveys were designed by the authors and were conducted by the Center for Population and Policy Studies (CPPS) of the University of Gadjah Mada, Yogyakarta, Indonesia.

The sample for the surveys covers each of the 300 subdistricts that were included in the original Generasi randomization. In each subdistrict, eight villages were randomly selected (unless the subdistrict had fewer than eight villages, in which case all were selected). This resulted in a total of 2,313 villages that was sampled in each of the three survey waves. Approximately 12,000 households were interviewed in each survey wave, as well as more than 8,000 village officials and health and education providers.

The sampling design for the household component of the Generasi surveys was chosen to ensure adequate coverage in the key Generasi demographic groups: mothers who recently were pregnant or gave birth, children under age 3, and children of school age. Within each village, one hamlet (dusun) was randomly selected, and a list of all households was obtained from the head of the hamlet. Five households were randomly sampled from that list to be interviewed. These households were stratified so that two selected households had at least one child under age 2, two selected households had a child under age 15 but no children under age 2, and one household had no children under age 15. In Wave II and Wave III, in half of the randomly selected villages (four villages out of the eight villages sampled in every subdistrict), the same households sampled in Wave I were contacted again in subsequent waves to form an individual level panel.[10] In the other half of villages, a new cross-section of households was drawn in each survey wave. The combination of panel households and non-panel households allows us to investigate heterogeneous treatment effects based on pre-period income levels and other characteristics, while at the same time ensuring that sufficient respondents with recent births and young children are enrolled in the survey sample in every round. Additional details about the surveys can be found in Olken, Onishi, and Wong (2011).

[10] Teams tracked and re-interviewed migrated or split households who provided information for any of the married women or children modules, as long as they were within the same subdistrict. In panel areas, 95% of target households were able to be reinterviewed in Wave 2 and 98% of target households were able to be reinterviewed in Wave 3.

2.4. Estimation

Since the Generasi program was designed as a randomized experiment, the evaluation is econometrically straightforward: essentially, we compare outcomes in those subdistricts randomized to be treatments with those subdistricts randomized to be control areas, controlling for the level of the outcome at baseline.

In implementing our analysis, we restrict attention to the 264 "eligible" subdistricts, as above, and use the randomization results combined with the government's prioritization rule to construct our treatment variables. Specifically, analyzing Wave II data (corresponding to the first treatment year), we define the *GENERASI* variable to be a dummy that take value 1 if the subdsitrict was randomized to receive *GENERASI* and either a) it was in the priority area (group P) or b) was in the non-priority area and selected in the additional lottery to receive the program in 2007. In analyzing Wave III data, we define the *GENERASI* variable to be a dummy that takes value 1 if the subdistrict was randomized to receive *Generasi*. We define the *GENERASI_INCENTIVES* variable to be a dummy that takes value 1 if the GENERASI is 1 and if the subdistrict was randomized to be in the incentivized version of the program. *GENERASI_INCENTIVES* thus captures the additional effect of the incentives above and beyond the main effect of having the program, and is the key variable of interest in the paper. Note that by defining the variables in this way, we are exploiting only the variation in program exposure due to the lottery. These variables capture the intent-to-treat effect of the program, and since the lottery results were very closely followed – they predict true program implementation in 99% of subdistricts in 2007 and 96% of subdistricts in 2008 – they will be very close to the true effect of the treatment on the treated (Imbens and Angrist 1994).

In running the regressions, we take advantage of the baseline data by controlling for the average level of the outcome variable in the subdistrict in the baseline survey. Since we also have individual-specific panel data for half our sample, we include the pre-period value for those who have it, as well as a dummy variable that corresponds to having non-missing pre-period values. Since households came from one of three different samples (those with a child under age 2, those with a child age 2–15 but not in the first group, and all others), we include dummies for those three sample types, interacted with whether a household came from a panel or non-panel village. Finally, since many of the indicators for children vary naturally as the child ages, for all child-level variables we include age dummies.

We thus estimate the following regressions:

Wave II data:

$$y_{pdsi2} = \alpha_d + \beta_1 GENERASI_{pds2} + \beta_2 GENERASI\_INCENTIVES_{pds2} + \gamma_1 y_{pdsi1} + \gamma_2 1_{\{ypdsi1 \neq missing\}} + \gamma_3 \overline{y_{pds1}} + SAMPLE_{pdsi} + \alpha_p \times P_s + \varepsilon_{pdsi} \quad (1)$$

Wave III data:

$$y_{pdsi3} = \alpha_d + \beta_1 GENERASI_{pds3} + \beta_2 GENERASI\_INCENTIVES_{pds3} + \gamma_1 y_{pdsi1} + \gamma_2 1_{\{ypdsi1 \neq missing\}} + \gamma_3 \overline{y_{pds1}} + SAMPLE_{pdsi} + \alpha_p \times P_s + \varepsilon_{pdsi} \quad (2)$$

Wave II and III combined average effect:

$$y_{pdsit} = \alpha_{dt} + \beta_1 GENERASI_{pdst} + \beta_2 GENERASI\_INCENTIVES_{pdst} + \gamma_{1t} y_{pdsi1} + \gamma_{2t} 1_{\{ypdsi1 \neq missing\}} + \gamma_{3t} \overline{y_{pds1}} + \pi_t SAMPLE_{pdsi} + \alpha_{pt} \times P_s + \varepsilon_{pdsit} \quad (3)$$

where p is a person, d is a district, s is a subdistrict, t is the survey wave (1 = baseline, 2 = interim survey, 3 = final survey), y_{pdsit} is the outcome in Wave t, α_d is a district fixed effect,

y_{pdsi1} is the baseline value for individual *i* (assuming that this is a panel household, and 0 if it is not a panel household), $1_{\{ypdsi1 \neq missing\}}$ is a dummy for being a panel household, $\overline{y_{ds1}}$ is the average baseline value for the subdistrict, SAMPLE are dummies for how the household was sampled interacted with being a panel or cross-section household, and $\alpha_p \times P_s$ are province-specific dummies for being in the sample areas having had prior community-driven development experience through the KDP program. Standard errors are clustered at the subdistrict level. Note that in the final equation for computing the average effect over Wave II and Wave III, all control variables (e.g., district FE, sample controls, baseline values, etc) are be fully interacted with wave dummies, to capture the fact that there may be differential trends in different parts of the country.

The key coefficient of interest is β_2, which estimates the difference between the incentivized and non-incentivized version of the program. We can also calculate the total impact of the incentivized version of the program (vis-à-vis pure controls) by adding the coefficients on *GENERASI_INCENTIVES* and *GENERASI*. We also examine a wide variety of additional specifications as robustness tests; these specifications are discussed in more detail in Section 3.

Since we have a large number of indicators, in order to calculate joint significance we will calculate average standardized effects for each family of indicators, following Kling, Liebman, and Katz (2007). Specifically, for each indicator *i*, define σ^2_i to be the variance of *i*. We then estimate (1) for each indicator, but run the regressions jointly, clustering the standard errors by subdistrict to allow for arbitrary correlation among the errors within subdistricts both between and across indicators. We then define the average standardized effect as

$$\sum_i \frac{\beta_i}{\sigma_i}. \qquad (4)$$

Note that all of the analysis presented here (regression specifications including control variables, outcome variables, and aggregate effects) follows an analysis plan that was finalized in April 2009 for the Wave II data (before we examined any of the Wave II data) and in January 22, 2010 (before we examined any of the Wave III data). This hypothesis document was registered with the Abdul Latif Jameel Poverty Action Lab at MIT and is available on request.

2.5. Summary statistics and balance

This section shows summary statistics and examines the balance of key child health and education indicators using data from the interviews of mothers in the baseline household surveys. The first 4 columns of Table 3 show the means of each of the 12 key indicator variables that correspond to the 12 target program indicators. Column (1) shows the mean in the control group, pooled across all 4 . Columns (2) – (4) then show the total means at baseline broken up by the three main regions under study – Java, NTT, and Sulawesi. Across virtually all 12 indicators, the off-Java locations appear substantially less developed. For example, malnutrition for under 3 year olds (defined as being more than 2 standard deviations below the age adjusted weight, as defined by the WHO) is 12.6 percent in Java, but 24.7 percent in NTT and 23.4 percent in Sulawesi. To take another example, 76.4 percent of births in our sample in Java are attended by a trained medical professional (midwife or doctor); by contrast, only 42.7 percent are attended by a trained professional in NTT and only 55.9 percent in Sulawesi. School participation rates are also lower off-Java, though the difference is not as stark as for the education indicators.

To check for balance across treatment arms, we estimate the relationship between the baseline survey values of the twelve major indicators that are the focus of the program (these indicators are discussed in more detail in Section 3 below) and the GENERASI and GENERASI_INCENTIVES variables. We use the specification as in equation (1) and (2) above (though naturally we don't control for baseline values, since the baseline values are on the left hand side of the regression.)

Column (5) shows the total estimated impact of the program in incentivized areas compared to pure controls, obtained by summing the coefficients on GENERASI and GENERASI_INCENTIVES, where GENERASI is defined based on the randomization results and prioritization rules for year 1 implementation. Column (6) shows the total estimated impact of the program in non-incentivized areas, which is the coefficient on the GENERASI variable. Column (7) shows the effect of the incentives relative to the non-incentives, which is the coefficient on GENERASI_INCENTIVES. Each row corresponds to a separate regression. Columns (8) – (10) repeat the same regressions, but with GENERASI and GENERASI incentives based on the randomization results for treatment in either year.

Looking across columns (5) through (10), we find that of the seventy-two coefficients estimated, eight are statistically significant at the 10 percent level or higher, which is precisely what would be predicted by random chance. Similarly, four of seventy-two coefficients are statistically significant at the 5 percent level or higher, and one is statistically significant at the 1 percent level, which is also what one would expect based on random chance. These results confirm that the randomization was indeed carried out properly and that the treatment and control groups are balanced.

The final rows of Table 3 consider the average standardized effects, computed via equation (3). We report average standardized effects for all twelve of the main indicators, and then separately report average standardized effects for the eight health indicators and four education indicators. Three of the eighteen coefficients are statistically significant at the 10 percent level, once again consistent with what would expect based on random chance. One coefficient is statistically significant at the 1 percent level, which we regard as a fluke. Most important, none of the average standardized effects for the additional effect of the incentives (columns 4 and 7), which are the key coefficients of interest for this paper, show any statistically significant differences at baseline.

3. Main results on targeted indicators

3.1. Overall impact

Table 4 presents the results on the 12 targeted indicators. Each row in Table 4 presents three separate regressions. Columns (2) – (4) show the regression with the Wave II survey results (after one year of program implementation) estimated using equation (1); columns (5) – (7) show the regression with the Wave III survey results estimated using equation (2); columns (8) – (10) show the regression that average across both survey waves estimated using equation (3). For each specification, we show the total Generasi treatment effect in incentive areas (the sum of the coefficients on GENERASI and GENERASI_INCENTIVES), the total Generasi treatment effect in non-incentive areas (the coefficient on GENERASI), and the additional treatment effect due to the incentives (the coefficient on GENERASI_INCENTIVES). The first 12 rows present each of the main 12 indicators one by one, and the bottom three rows present average standardized effects overall, for the 8 health indicators and for the 4 education indicators. All data is from the household surveys.

Focusing first on the average standardized, column (10) shows that on average over the two years of the program, the 8 targeted maternal and child indicators (e.g., prenatal visits, delivery by trained midwives, immunizations, regular weight checks) were an average of 0.03

standard deviations higher in incentivized areas than in non-incentivized areas. The only one of the 12 indicators that is individually significant is prenatal visits, which were 5% higher in incentivized areas than non-incentivized areas. The effects appear somewhat larger in Wave II – column (4) shows that the average standardized effect of the incentives on health is 0.04 standard deviations, whereas column (7) shows that they were a (statistically insignificant) 0.026 standard deviations in Wave III. The comparison to pure controls suggest that the change over time is due to increases in effectiveness in the non-incentivized areas, rather than declining effectiveness of the incentivized areas, though these differences are not statistically significant.[11] In Wave II prenatal visits were 8% higher in incentivized areas than in non-incentivized areas; weight checks were 4% higher; and, most notably, malnutrition was 15% lower in incentivized areas. No effects of the incentives were seen in education in either wave.

It is worth noting that a substantial share of the overall effect of the Generasi program can be attributed to the performance incentives. In Wave II, the incentivized version of the program improved the 12 indicators by an average of 0.053 standard deviations compared to pure control (column 2), while the non-incentivized version improved the 12 indicators by only 0.012 standard deviations (column 3, statistically insignificant). This implies that 77% of the total impact of the program can be attributed to the incentives. In Wave III, even though the effect of the incentives is statistically insignificant, the point estimates suggest that 50% of the total impact of the program can be attributed to the incentives. Thus, when the incentive effect is scaled by the overall impact of the program, the incentives seem to have had a substantial effect.

[11] To test whether the differences between waves are statistically significant, we conducted additional analysis where we restricted the sample to those subdistricts that were randomized to be treatment in both waves or control in both waves, so that differences over time could be separated from changing composition of districts. We then tested whether the impact of incentives in Wave II (after 1 year) was different from the impact in Wave III (after 2 years). Appendix Table 1 presents the results, which shows that none of the differences in average standardized effects for health (either separately by treatment or the additional effects of the incentives) are statistically significantly different between waves.

An alternative approach to weighting the various individual effects is to use the weights calculated by the program, shown in Table 1. This approach has the advantage that it weights each indicator by the effective weight assigned to it by the government in constructing the program. For each indicator, we use the weights in Table 1, multiplied by the number of potential beneficiaries of each indicator (garnered from population data in different age ranges from the program's internal management system), and aggregate to determine the total number of "points" created by each version of the program. The results, shown at the bottom of Table, show a similar story to the average standardized effects. In Wave II, 93 percent of the program's impact on health (in terms of points) can be attributed to the incentives; in Wave III, 30 percent of the program's impact on health (in terms of points) can be attributed to the incentive, though the Wave III difference is not statistically significant. We use the estimated impact on the number of points in Section 6 below to examine the cost-effectiveness of the program.

Although we pre-specified equations (1) – (3) as the main regression specifications of interest, we have also considered the robustness of these results to a wide range of alternative specifications. Appendix Table 1 reports the coefficient on the additional effect of the incentives – the equivalent of columns (4), (7), and (10) – for specifications where we control for the baseline level of all 12 indicators instead of just the indicator in question, control only for subdistrict averages at baseline rather than also using individual baseline controls, include no controls whatsoever, estimate the regression in first-differences rather than including the baseline level as a control, and run the entire regression aggregated to the subdistrict level, rather than using individual level data. The results are very consistent with the main specification in Table 4.

3.2. Heterogeneity in impact

The incentives also had larger impacts in the off-Java locations where baseline levels of services were lower. Table 5 repeats the previous analysis, but fully interacts the GENERASI and GENERASI_INCENTIVES variables with dummies for being in NTT and Sulawesi. Given the large number of coefficients, we report the average results over both years only in Table 5; the results for Wave II and III separately are shown in the appendix. The results in Table 5 show that the incentives were substantially more effective off Java: on average, the total standardized effect for health was 0.11 standard deviations higher in incentivized areas than non incentivized areas in NTT relative to Java, and 0.14 standard deviations higher in incentivized areas than non incentivized areas in Sulawesi relative to Java. The appendix shows that the effects are similar in Wave II and Wave III if we examine the effects separately. Once again, there is less of an impact on education, with the only impacts coming in Sulawesi, where the incentives appear to reverse the (surprising) negative effects of the program in the non-incentivize areas.

One reason the incentives may have been more effective off-Java is that, with lower levels of baseline service delivery, the indicators were easier to move in those areas. To examine this explicitly, we re-estimate equations (1) – (3), this time interacting the GENERASI and GENERASI_INCENTIVES variables with the mean value of the indicator in the subdistrict at baseline. The results are shown in Table 6. A negative coefficient on the interaction implies that the program was more effective in areas with worse baseline levels. For ease of interpretation, we also present the implied impacts calculated at the 10^{th} percentile of the baseline distribution. The results confirm that the incentives were more effective in areas with lower baseline levels of service delivery – the standardized interaction term of GENERASI_INCENTIVES * BASELINE_VALUE in columns (3), (7), and (11) are negative and, in both Wave II and overall, statistically significant. To interpret the magnitude of the heterogeneity, note that, in Wave II, the incentives added 0.074 standard deviations to the health indicators at the 10th percentile of the baseline distribution. In Wave III, it was 0.06 standard deviations (not statistically significant), and across the two waves, it was 0.066 standard deviations. These effects are about double the average effect of the program shown in Table 4, and suggest that, indeed, the program was substantially more effective in areas with lower baseline levels.

4. Potential pitfalls of incentives

The previous sections shows that the performance incentives substantially increased the effectiveness of the program. In this section, we test for three types of negative consequences from the incentives: multi-tasking problems (Holmstom and Milgrom 1982), where performance incentives encourage substitution away from non-incentivized outcomes; manipulation of performance records; and reallocation of funds towards wealthier areas.

4.1. Spillovers on non-targeted indicators

Whether the incentives would increase or decrease performance on non-targeted indicators depends on the nature of the health and education production functions. For example, if there is a large fixed cost of getting a midwife to show up in a village, but a small marginal cost of seeing additional patients once she is there, one might expect that other midwife-provided health services would increase. Alternatively, if the major cost is her time, she may substitute towards the types of service incentivized by Generasi and away from things outside the incentive scheme, such as family planning, or might spend less time with each patient.

We test for spillover effects on 3 health domains: utilization of non-incentivized health services (e.g., adult health, prenatal visits beyond the number of visits that qualify for incentives), quality of health service provided by midwives (as measured by the share of the total required services they provide in a typical meeting), and maternal knowledge and practices. In constructing these indicators, we erred on the side of including more rather than fewer indicators. We also examine potential impacts on family composition decisions (for example, does better maternal care induce people to have more children or migrate into the area). On the education side, we examine the impact on high school enrollment, hours spent in school, enrollment in informal education (so-called Paket A, B, and C, which are the at-home equivalents of primary, junior secondary, and senior secondary schools), distance to school, and child labor.

In general, we find no differential negative spillover impacts of the performance incentives on any of these indicators. Looking at the indicators one-by-one, if anything we find some slight evidence of positive spillovers on some indicators. For example, we find that the performance incentives led positive effects of the incentives on prenatal visits beyond the 4 visits that were incentivized by the program (0.03 extra visits); on enrollment in the informal version of senior secondary school (0.1 percent of children in the same), and reductions in child labor (.12 hours per child for age 7-15), though the average standardized effects are only significant for child labor. This suggests that negative spillovers on non-targeted indicators does not seem to be a substantial concern with the performance incentives in this context.

4.2. *Manipulation of performance records*

A second potential downside of performance incentives is that communities or providers may manipulate records to inflate scores. For example, Linden and Shastry (2009) show that teachers in India inflate student attendance records to allow them to receive subsidized grain. Manipulation of recordkeeping can have substantial efficiency costs: for example, children could fail to get immunized properly if their immunization records were falsified.[12]

[12] This was a substantial concern was a substantial concern during program design. The idea was that with the performance bonuses being relative to other, nearby villages, surrounding villages would have an incentive to make sure that scores were not being manipulated. Moreover, since health providers and junior secondary scores

We have two ways in which we can check for manipulation of recordkeeping. First, the BCG vaccine leaves a distinctive scar on the arm, so we can compare a child's records on whether the BCG vaccine was administered to the presence of the BCG scar as monitored in our survey (Banerjee, Duflo, Glennerster, and Kothari 2010). Second, we compare attendance from random spot-checks of classrooms with attendance records from the same classroom on a specific day 1-2 months previously. Although the date compared are not the identical (we could not obtain reliable records on the date of our survey, since the presence of our surveyors might affect the records), the difference between them should capture, on average, the markup in attendance at the school.

The results are shown in Table 8. Panel A explores the differences between BCG scars and record keeping. Note that in if the child did not have a record card, we asked the mother if the child was immunized. The "declared" vaccinated variable is 1 if either the record book or the mother report that the child was vaccinated. On average, 75 percent of children have the scar; 60% of children have a record of receiving the vaccine, and 85 percent of children either have a record of receiving the vaccine or have a parent who reports the child received a vaccine. We defined a false "yes" if the child is recorded/declared as having had the vaccine but has no scar, and likewise for a false "no." We find no statistically significant differences in false reports of the BCG scar based on the performance incentives.

Panel B explores differences in attendance rates. On average, attendance is overstated: 88 percent of children were recorded as present by our random visits, whereas 95 percent were recorded present in the official attendance records. The discrepancy is unchanged by the performance incentives. Combined, these two pieces of evidence suggest that manipulation of

---

often serve many villages within a given subdistrict, they would have little to gain by manipulating scores.

recordkeeping is not a major problem of the performance incentives in this context, at least to the degree we can measure it.

5. Mechanisms

The results thus far showed that the incentives substantially improved the targeted health indicators with little obvious downside. In this section we explore three potential mechanisms through which the incentives may have had an impact: by inducing a change in the allocation of funds and by changing provider effort.

5.1. Allocation of funds

Table 9 examines whether the incentives affected how the Generasi communities chose to allocate the block grants. Each row in Panels A and B shows the share of the village's block grant spent on the item.

The most notable finding that emerges is that the incentives led to a shift away from education supplies – uniforms, books, and other school supplies – and towards health expenditures. In particular, spending on education supplies is about 4 percentage points (15 percent) lower in incentivized villages, and health spending is about 3 percentage points (7 percent) higher.[13] One interpretation of this is that these types of education supplies are essentially a transfer – when distributed, they tend to be distributed quite broadly to the entire population, the vast majority of whose children are already in school, and therefore are likely to have relatively little impact on school attendance and enrollment. As shown in Table 4 above, the performance incentives improved health outcomes with no detrimental effect on education, so combined this suggests that the performance incentives may have led communities to

[13] In the appendix, we try to further break down health to determine which health categories increase, but we do not have the statistical power to do so meaningfully.

reallocate funds away from (potentially politically popular) ineffective education spending towards more effective health spending.

We also tested two other hypotheses that do not seem borne out in the data. First, we expected that, since performance incentives effectively increase the discount rate (since one places higher value on a return in the current year since it will affect bonuses), we would expect a shift away from durable investments – if anything, the opposite appears to have occurred, with spending on health durables increased by about 1.7 percentage points (18 percent). Second, we expected that performance incentives would lead to a decrease in "capture" of the funds to expenses benefitting providers (e.g., uniforms for providers), but we see no impact on this dimension.

The evidence thus far was on how the money was spent. Table 10 shows the other side of the equation, namely, what households received, using data from the household survey. Two items are noticeable here. First, households were no less likely to receive a uniform or school supplies in the incentive treatments than in the non-incentive treatments – in fact, if anything the point estimates suggest they were 1.0-2.7 percentage points (12-32 percent) more likely to receive a uniform in the groups with performance incentives and 1.0-1.7 percentage points (18-32 percent) more likely to receive other school supplies in the groups with performance incentives. This suggests that the change in budget allocations towards these items shown in Table 9 was more likely to be on the quality dimension (nicer uniforms) than on the quantity dimension, and thus may have been particularly efficient. Second, there was also a substantial an increase in scholarships (1 percentage point, or about 125 percent) and transport subsidies (0.6 percentage points, about 110 percent). Thus, on average more children received education subsidies, even though on more money was being spent on health.

Although it is very hard to conclusively measure the efficiency of allocations, combined, the evidence in these tables suggest that the performance incentives led to a decrease in the quality of school supplies given out, which is likely to be efficiency improving, and a reallocation of expenditures towards health. Combined with the fact that the health outcomes improved and education did not suffer, this provides suggestive evidence that the performance incentives improved the efficiency of the Generasi funds.

5.2. Effort

A second dimension we examine is effort – both on the part of worker and on the part of communities.

Did the performance incentives increase effort? Table 11 begins by examining effort of midwives, who are the primary health workers at the village level, teachers, and subdistrict level health center workers. The main impact is an increase in labor on the part of midwives. On average, midwives spent 1.7 hours (6 percent) more working over the 3 days prior to the survey in incentive areas than in non-incentive areas. There was no impact on teacher attendance or provider attendance at health care centers. Given that midwives are the main providers of the targeted maternal and child health services, this increase in midwife effort is consistent with the increase in these services.

Table 12 examines the effort of communities. We examine three types of community effort: holding more posyandus, the monthly village health meetings where most maternal and child health care is providers; community effort at outreach, such as door-to-door "sweepings" to get more kids into the posyandu net and school committee meetings with parents, and community effort at monitoring service providers, such as school committee membership and meetings with teachers. We find no evidence that the performance incentives had an impact on any of these margins.

6. Cost-effectiveness

It is difficult to interpret the magnitudes given above without some notion of costs. Conditional on implementing the Generasi program, adding the performance incentives was essentially free – the same monitoring of indicators was done in both the incentivized and non-incentivized version of the program, and since the performance bonuses were relative within a subdistrict and the amount of money was fixed, there was no difference in the total size of block grants in incentivized and non-incentivized areas. Thus the cost effectiveness of the incentives themselves for this program is easy to analyze: they improved outcomes, added virtually no costs, and therefore by definition must be cost effective.

However, it is also useful to gauge the overall cost-effectiveness of the program. The challenge in doing so is that there are many potential outcomes, and we do not necessarily know how to apportion the costs of the programs among the various outcomes. We therefore take the following approach: as in Section 3 above, we calculate the total benefits of the program as the total number of "points" the program created, using the weighting scheme agreed upon in advance and shown in Table 1, an estimate based on the MIS data of the total number of people in each age range who could have benefitted on each dimension, and the point estimates for the impact of the program from Table 4.[14] We then divide the total cost of the program by the total number of points created to generate a "cost per point", which can then be interpreted using the

[14] Note that the number of points shown in this section is approximately half the total number of points reported in Table 4. The reason is that Table 4 is calculated using the total number of beneficiaries in the program, whereas in this section since we are mapping it to points we use number of beneficiaries in the particular treatment group only, which is half of the total.

point values in Table 1. While naturally different weighting schemes could produce different answers, we use the points in the program since they presumably represent the government's relative weightings of the different interventions, and also since they are fixed in advance and thus not subject to manipulation.

In terms of allocating the costs of the program, we pursue two approaches. First, we take a conventional public-finance approach. The conventional public-finance approach to cost-effectiveness counts the real social cost of transfers as only the social deadweight loss of taxation used to raise the revenue used to fund them. The idea here is that transfers are valued by recipients at cost, so the only cost is the deadweight loss of taxation. Non-transfer costs (such as hiring a midwife) are counted as "real" costs, and thus the true social cost is the expenditure cost plus the deadweight loss of taxation. For the purposes of evaluating Generasi, we count school supplies, school fee subsidies, health care subsidies, and supplementary food as transfers, and all other expenditures as real expenditures. As shown in Table 9 above, about 75% of the block grant are spent on transfers by these definitions. We also include the cost of the facilitators who administer the program as real expenditures. We use the consensus estimate of the marginal cost of public funds of 0.3 (Ballard et al), though note that there are not reliable estimates of this parameter for developing countries. Second, we also report an alternative estimate where we just count the total dollar cost of the program, implicitly valuing transfers and non-transfers the same, and do not consider the marginal cost of taxation. While we view this method as less appropriate from an economic perspective, it is how many donors do cost-effectiveness calculations, so we present it for ease of comparison in Panel C of Table 13. We use the Wave III analysis (at the end of the program's second year), when the program was at full scale, for this calculation.

The estimates are presented in Table 13. The key results are shown in the first two columns of Panel A: Generasi with incentives had a real cost per point of about $8, and Generasi without incentives had a real cost per point of about $11. Since the estimates in Table 4 show that, for year 3, the difference in the total number of points between incentivized and non-incentivized versions of the program is not significantly different, we should treat the difference between $8 and $11 is also not statistically significant.

Panel B separately estimates the cost effectiveness for the health and education components of the program, allocating facilitation costs equally between the two portions of the program and allocating expenditures based on how communities actually allocated block grants. For health, this yields estimates of $7 per point for the incentive version and $9 for the non-incentivized version. For education, this yields estimates of $13 per point for the incentive version and $16 for the non-incentivized version.

How do we interpret the $8 - $11 per point average cost effectiveness of the program? One way to think about it is that – certainly for the indicators that did in fact move significantly – we can back out what this implies to move an indicator. Applying the weights from Table 1, for example, suggests that the cost of additional child weight check was $16 - $22, the cost of preventing one malnourished child was $384 - $528, the cost of getting one additional child fully covered with Vitamin A was $160 - $220, and the cost of enrolling one more child in primary school was $200 - $275.

Are these numbers large or small? While that is ultimately a judgment question for the reader, we provide two benchmarks. First, the closest comparison is Indonesia's conditional cash transfer program (PKH). The PKH program was conducted at the same time and evaluated using a randomized evaluation using the same survey instruments as Generasi, though it was conducted in somewhat different areas of the country (more urban and with better supply of services), and was targeted at the same set of indicators. We use the randomized evaluation results from Alatas et al (2011) of the PKH program, combined with the same weights in Table 1. Alatas et al reports an estimated effect just for those households receiving PKH, as well as a "placement effect" on all poor households in the subdistricts regardless of whether they received PKH or not. We report cost-effectiveness numbers based on both calculations. The results suggest that, if one focuses only on the benefits enjoyed by PKH households, Generasi is more cost effective – with the $8 – $11 per point in Generasi comparing to about $22 per point for PKH. If one includes estimated spillover effects from PKH to non-recipient households in the same subdistricts, then the $8 - $11 per point for Generasi is comparable to the $11 per point estimate for PKH.[15] Thus, the Generasi program looks roughly comparable to an alternative program tried in Indonesia at the same time.

An alternative benchmark is to look at international comparisons. On the education side, we consider come from the Poverty Action Lab's review of cost-effectiveness of getting additional years of school attendance (JPAL, 2011). They show, for example, that school-based deworming in Kenya costs $3.50 per additional year of school attendance, and iron and deworming tablets in India cost $29 per additional year of school attendance. School meals in Kenya cost $35 per additional year of attendance, and school uniforms cost about $100 per additional year of attendance. The JPAL methodology does not count the marginal cost of public funds, so the comparable numbers are the ones presented in Panel C of Table 13. Using the estimates in Panel C of Table 13 ($5 - $6 per point not counting the cost of transfers, or $12 - $16 per point counting the transfers), Generasi would cost between $125 - $400 per additional

[15] It is also worth noting that neither calculation includes the benefits from redistribution. Since both programs give transfers to poorer households, financed by taxes on richer programs, both

year schooling. By this metric, Generasi as a whole is substantially less cost-effective than these other interventions, although it is worth noting that Generasi affects enrollment rates, whereas the international comparisons affect attendance, and that there is already high baseline enrollment in Generasi areas, which makes the marginal cost higher.

On the nutrition side, one relevant program is a supply-side nutrition program in Peru (Waters et al). This study reports that the cost was US$55 per case of stunting averted. While malnutrition (weight for age) is not the same as stunting (height for age), it is worth noting that, by comparison, Generasi would cost $240 - $800 per case of malnutrition avoided.

While these numbers suggest that Generasi, as a whole, may be more expensive than these other programs, the performance incentives themselves – at $0.62 per point, which translates into $16 per additional child enrolled in school and $30 per case of malnutrition avoided – compares favorably with all of the above interventions except deworming. The reason the incentives themselves are cost effective is that in our case they are essentially free – the block grant was the same with and without incentives, and collecting the data used to validate the incentives was done in both the incentivized and non-incentivized versions of the program, so the only "costs" come from the fact that there were slightly more real expenditures and slightly fewer transfers in the incentivized version of the program. Even if one had to pay for all the costs of data collection, that would likely still be small, since the block grant program still needs facilitators in any case. Although Kenya, India, and Peru are different in many ways from Indonesia, so these estimates are not directly comparable, this conclusion suggests that while the

Generasi program as a whole was not as cost effective as these other interventions, adding incentives to existing programs may be a cost-effective way to improve performance.[16]

7. Conclusion

In sum, the evidence presented here suggests that properly designed, performance based incentives can be a useful addition to aid programs. We found that adding performance incentives increased health outcomes, particularly in poorer areas with worse performance at baseline. Though the gains from incentives were modest, we found little downside from the incentives.

[16] The same study estimates a cost of $5000 per additional year of primary attendance from the Progresa program, but that estimate is not comparable to the estimates here since the authors assign *all* of the costs of Progresa to primary school attendance, which is not comparable to the methodologies we use here.

References

Birdsall, N. and W.D. Savedoff (2009). *Cash on Delivery: A New Approach to Foreign Aid With an Application to Primary Schooling*. Washington, DC, Center for Global Development.

Duflo, E., R. Hanna and S. Ryan (2008). "Monitoring Works: Getting Teachers to Come to School ", MIT.

Express India News Service (2008). "Biometric attendance to keep track of students, teachers in primary schools." *Express India*.

Gertler, P. (2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA' s Control Randomized Experiment." *American Economic Review 94*(2): 336-341.

Gibbons, R. and K.J. Murphy (1990). "Relative Performance Evaluation for Chief Executive Officers." *Industrial and Labor Relations Review 43*(3): 30-51.

Holmstrom, B. (1979). "Moral Hazard and Observability." *Bell Journal of Economics 10*(1): 74-91.

Holmstrom, B. and P. Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organizations 7*: 24.

Imbens, G.W. and J.D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica 62*(2): 467-475.

Lazear, E.P. and S. Rosen (1981). "Rank-Order Tournaments as Optimum Labor Contracts." *The Journal of Political Economy 89*(5): 841.

Levy, S. (2006). *Progress against poverty: sustaining Mexico's Progresa-Oportunidades program*, Brookings Institution Press, Washington, DC.

Linden, L. and K. Shastry (2008). "Identifying Agent Discretion: Exaggerating Student Attendance in Response to a Conditional School Nutrition Program." Columbia University.

Mookherjee, D. (1984). "Optimal Incentive Schemes with Many Agents." *Review of Economic Studies 51*(3): 433-46.

Schultz, T.P. (2004). "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics 74*(1): 199-250.

Weitzman, M. (1980). "The ratchet principle and performance incentives." *Bell Journal of Economics 11*(1): 302-8.

Table 1: *Generasi* program target indicators and weights

Performance metric	Weight per measured achievement	Potential times per person per year	Potential points per person per year
1. Prenatal care visit	12	4	48
2. Iron tablets (30 pill packet)	7	3	21
3. Childbirth assisted by trained professional	100	1	100
4. Postnatal care visit	25	2	50
5. Immunizations	4	12	48
6. Monthly weight increases	4	12	48
7. Weight check	2	12	24
8. Vitamin A pill	10	2	20
9. Primary enrollment	25	1	25
10. Monthly primary attendance >= 85%	2	12	24
11. Middle school enrollment	50	1	50
12. Monthly middle school attendance >= 85%	5	12	60

Table 2: *Generasi* randomization and implementation

	Incentivized *Generasi*		Non-incentivized *Generasi*		Control		Total
	P	NP	P	NP	P	NP	
Total subdistricts in initial randomization	61	39	55	45	55	45	300
Total eligible subdistricts	57	36	48	40	46	37	264
Eligible and received *Generasi* in:							
2007	57	10	48	12	0	0	127
2008	57	33	48	36	0	0	174

Table 3. Baseline Regressions and Summary Statistics, 12 main indicators

	Summary statistics				Year 1 Treatment			Year 1 or 2 Treatment		
Indicator	Control Group Mean	Mean in Java	Mean in NTT	Mean in Sulawesi	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number prenatal visits	7.808	8.298	7.323	6.518	-0.341	-0.125	-0.216	-0.3794*	-0.186	-0.194
	[4.4482]	[4.0927]	[4.6628]	[5.1344]	(0.263)	(0.267)	(0.259)	(0.228)	(0.223)	(0.208)
Delivery by trained midwife	0.691	0.782	0.489	0.579	-0.020	-0.026	0.005	-0.035	-0.018	-0.017
	[0.4623]	[0.4136]	[0.5013]	[0.4956]	(0.029)	(0.030)	(0.031)	(0.026)	(0.026)	(0.025)
Number of postnatal visits	3.012	3.635	1.697	2.143	-0.010	0.195	-0.206	-0.080	0.023	-0.102
	[3.3234]	[3.4562]	[2.5263]	[2.9785]	(0.196)	(0.211)	(0.199)	(0.190)	(0.192)	(0.172)
Iron tablet sachets	1.591	1.684	1.561	1.266	-0.071	-0.018	-0.054	-0.058	0.008	-0.065
	[1.2790]	[1.2962]	[1.3150]	[1.1103]	(0.070)	(0.078)	(0.073)	(0.066)	(0.067)	(0.061)
Percent of immunization	0.680	0.726	0.595	0.607	-0.0381*	-0.036	-0.003	-0.0509***	-0.0416**	-0.009
	[0.3519]	[0.3307]	[0.3738]	[0.3737]	(0.021)	(0.023)	(0.023)	(0.019)	(0.019)	(0.018)
Number of weight checks	2.140	2.265	2.351	1.490	-0.091	-0.047	-0.044	-0.1509**	-0.046	-0.1053*
	[1.1898]	[1.1194]	[1.0641]	[1.3248]	(0.060)	(0.069)	(0.065)	(0.058)	(0.059)	(0.055)
Number Vitamin A supplements	1.521	1.473	1.589	1.611	0.097	-0.031	0.1275*	0.057	-0.035	0.0922*
	[1.1686]	[1.0718]	[1.2016]	[1.4227]	(0.061)	(0.063)	(0.066)	(0.058)	(0.060)	(0.056)
Percent malnourished	0.173	0.121	0.245	0.263	0.022	0.006	0.015	0.009	0.001	0.008
	[0.3784]	[0.3258]	[0.4311]	[0.4415]	(0.016)	(0.015)	(0.017)	(0.014)	(0.014)	(0.014)
Age 7–12 participation rate	0.950	0.963	0.918	0.954	0.006	-0.006	0.012	0.000	0.001	-0.001
	[0.2191]	[0.1884]	[0.2747]	[0.2098]	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Age 13–15 participation rate	0.825	0.870	0.708	0.808	0.017	0.000	0.017	0.001	0.018	-0.016
	[0.3801]	[0.3364]	[0.4571]	[0.3955]	(0.027)	(0.028)	(0.027)	(0.024)	(0.025)	(0.022)
Age 7–12 gross attendance	0.910	0.931	0.873	0.896	-0.006	-0.0287**	0.023	-0.011	-0.011	0.000
	[0.2685]	[0.2379]	[0.3195]	[0.2760]	(0.014)	(0.014)	(0.016)	(0.013)	(0.012)	(0.013)
Age 13–15 gross attendance	0.752	0.797	0.602	0.759	0.036	0.003	0.033	0.027	0.036	-0.009
	[0.4196]	[0.3869]	[0.4870]	[0.4167]	(0.031)	(0.032)	(0.033)	(0.029)	(0.030)	(0.027)
Average standardized effect					-0.016	-0.029	0.013	-0.0365*	-0.012	-0.024
					(0.023)	(0.024)	(0.023)	(0.021)	(0.022)	(0.020)
Average standardized effect health					-0.0415*	-0.027	-0.014	-0.0589***	-0.032	-0.027
					(0.025)	(0.029)	(0.026)	(0.022)	(0.024)	(0.022)
Average standardized effect educ.					0.035	-0.031	0.066	0.008	0.026	-0.018
					(0.047)	(0.050)	(0.049)	(0.042)	(0.044)	(0.040)

Notes: Columns (1) – (4) show the means of the variable shown, with standard deviations in brackets. Each row of columns (5) – (7) and (8) – (10) show coefficients from a regression of the variable shown on an incentive treatment dummy, a non-incentive treatment dummy, district fixed effects, and province * group P fixed effects. Robust standard errors in parentheses, adjusted for clustering at the subdistrict level. In columns (5) – (7) the treatment variable is defined based on year 1 program placement, and in columns (8) – (10) it is defined based on year 2 program placement. Columns (7) and (10) are the calculated difference between the previous two columns. All treatment variables are defined using the original randomization, and so are interpretable as intent-to-treat estimates. Average standardized effects reported in the bottom three rows are calculated using the estimated coefficients from the 12 individual regressions above using the formula shown in the text, adjusted for arbitrary cross-equation clustering of standard errors within subdistricts. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table 4: Impact on targeted outcomes

	Wave II				Wave III			AVERAGE		
Indicator	Baseline Mean	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Average Treatment Effect	Non-Incentive Average Treatment Effect	Incentive Average Additional Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number prenatal visits	7.447	0.333	-0.280	0.6129***	0.156	-0.025	0.181	0.235	-0.136	0.3706**
	[4.2935]	(0.233)	(0.200)	(0.220)	(0.192)	(0.188)	(0.173)	(0.155)	(0.139)	(0.147)
Delivery by trained midwife	0.670	0.034	0.040	-0.005	0.011	-0.008	0.019	0.023	0.014	0.009
	[0.4705]	(0.026)	(0.026)	(0.025)	(0.021)	(0.023)	(0.021)	(0.019)	(0.019)	(0.018)
Number of postnatal visits	1.720	-0.169	-0.065	-0.104	-0.034	-0.031	-0.003	-0.099	-0.051	-0.048
	[2.4477]	(0.140)	(0.120)	(0.140)	(0.129)	(0.124)	(0.129)	(0.099)	(0.091)	(0.101)
Iron tablet sachets	1.588	0.129	0.050	0.078	0.076	0.045	0.031	0.1008*	0.049	0.052
	[1.2554]	(0.084)	(0.081)	(0.081)	(0.058)	(0.065)	(0.063)	(0.052)	(0.052)	(0.051)
Percent of immunization	0.653	0.025	0.010	0.015	0.010	-0.007	0.017	0.018	0.002	0.016
	[0.3664]	(0.018)	(0.018)	(0.018)	(0.015)	(0.015)	(0.014)	(0.012)	(0.011)	(0.011)
Number of weight checks	2.126	0.1633***	0.068	0.0958*	0.1747***	0.1983***	-0.024	0.1672***	0.1394***	0.028
	[1.1895]	(0.052)	(0.049)	(0.054)	(0.055)	(0.052)	(0.051)	(0.041)	(0.039)	(0.040)
Number Vitamin A	1.529	-0.008	0.005	-0.013	0.0847*	0.002	0.083	0.038	0.000	0.037
supplements	[1.1370]	(0.052)	(0.055)	(0.058)	(0.048)	(0.054)	(0.053)	(0.035)	(0.040)	(0.038)
Percent malnourished	0.168	-0.016	0.011	-0.0265*	-0.017	-0.0262*	0.009	-0.016	-0.009	-0.006
	[0.3739]	(0.016)	(0.015)	(0.016)	(0.014)	(0.015)	(0.016)	(0.011)	(0.011)	(0.013)
Age 7–12 participation rate	0.948	-0.001	0.003	-0.004	0.005	0.0108***	-0.006	0.003	0.0075**	-0.005
	[0.2221]	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Age 13–15 participation rate	0.822	-0.0343*	-0.0504**	0.016	0.020	0.013	0.007	-0.004	-0.015	0.011
	[0.3827]	(0.020)	(0.023)	(0.024)	(0.017)	(0.016)	(0.014)	(0.014)	(0.016)	(0.014)
Age 7–12 gross attendance	0.904	0.001	0.002	-0.001	0.003	0.004	-0.001	0.002	0.003	-0.001
	[0.2773]	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.004)	(0.005)
Age 13–15 gross attendance	0.768	-0.0405*	-0.0651***	0.025	0.025	0.016	0.010	-0.005	-0.020	0.015
	[0.4125]	(0.021)	(0.024)	(0.025)	(0.018)	(0.017)	(0.015)	(0.015)	(0.016)	(0.015)
Average standardized effect		0.014	-0.022	0.036	0.0502**	0.0320*	0.018	0.0345**	0.010	0.025
		(0.023)	(0.021)	(0.024)	(0.020)	(0.018)	(0.019)	(0.016)	(0.016)	(0.016)
Average standardized effect		0.0526**	0.012	0.0406*	0.0515**	0.026	0.026	0.0516***	0.020	0.0316*
Health		(0.024)	(0.023)	(0.024)	(0.023)	(0.022)	(0.022)	(0.018)	(0.018)	(0.018)
Average standardized effect		-0.063	-0.0899**	0.027	0.048	0.0449*	0.003	0.000	-0.011	0.011
educ.		(0.039)	(0.045)	(0.045)	(0.029)	(0.027)	(0.027)	(0.026)	(0.029)	(0.026)
Total points		0.614	-1.717	2.3306*	2.8330**	2.0873*	0.746	1.7969*	0.403	1.394
(millions)		(1.365)	(1.255)	(1.408)	(1.247)	(1.149)	(1.176)	(1.011)	(0.958)	(0.997)
Total points health		1.8646*	0.135	1.7298*	1.9166*	1.334	0.583	1.8614**	0.788	1.074
(millions)		(0.971)	(0.920)	(1.001)	(0.983)	(0.979)	(0.969)	(0.765)	(0.759)	(0.797)
Total points education		-1.2507*	-1.8514**	0.601	0.916	0.754	0.163	-0.065	-0.385	0.320
(millions)		(0.712)	(0.823)	(0.836)	(0.585)	(0.554)	(0.516)	(0.495)	(0.553)	(0.494)

Notes: Column 1 shows the baseline mean of the variable shown, with standard deviations in brackets. Each row of columns (2) – (4), (5) – (7), and (8) – (10) show coefficients from a regression of the variable shown on an incentive treatment dummy, a non-incentive treatment dummy, district fixed effects, province * group P fixed effects, and baseline means, as described in the text. Robust standard errors in parentheses, adjusted for clustering at the subdistrict level. In columns (2) – (4) the treatment variable is defined based on year 1 program placement, and in columns (5) – (7) it is defined based on year 2 program placement, and in columns (8) – (10) , which uses pooled data from both waves, it is defined as year 1 placement for the Wave II data and as year 2 placement for the Wave III data. All treatment variables are defined using the original randomization, and so are interpretable as intent-to-treat estimates. Columns (4), (7), and (10) are the calculated difference between the previous two columns. Average standardized effects and total points reported in the bottom rows are calculated using the estimated coefficients from the 12 individual regressions above using the formula shown in the text, adjusted for arbitrary cross-equation clustering of standard errors within subdistricts. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table 5: Regional Heterogeneity

Indicator	Java Incentive Treatment Effect	Additional NTT Incentive Treatment Effect	Additional Sulawesi Incentive Treatment Effect	Java Non-Incentive Treatment Effect	Additional NTT Non-Incentive Treatment Effect	Additional Sulawesi Non-Incentive Treatment Effect	Java Incentive Additional Effect	Additional NTT Incentive Additional Effect	Additional Sulawesi Incentive Additional Effect
	AVERAGE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Main 12 indicators									
Number prenatal visits	0.296	-0.423	0.234	0.034	-0.6673*	-0.146	0.262	0.245	0.380
	(0.179)	(0.360)	(0.537)	(0.162)	(0.362)	(0.435)	(0.167)	(0.321)	(0.615)
Delivery by trained midwife	0.021	-0.033	0.059	0.0472**	-0.1359**	-0.016	-0.026	0.1026**	0.075
	(0.020)	(0.055)	(0.044)	(0.019)	(0.057)	(0.048)	(0.018)	(0.049)	(0.053)
Number of postnatal visits	-0.2928**	0.4672**	0.5560**	-0.055	-0.149	0.260	-0.2380*	0.6164***	0.296
	(0.128)	(0.213)	(0.267)	(0.128)	(0.196)	(0.216)	(0.132)	(0.207)	(0.290)
Iron tablet sachets	0.095	0.013	0.050	0.1240*	-0.156	-0.2756**	-0.029	0.1681*	0.3259**
	(0.069)	(0.113)	(0.135)	(0.070)	(0.116)	(0.123)	(0.069)	(0.101)	(0.135)
Percent of immunization	-0.006	0.0598**	0.0826*	-0.010	0.031	0.031	0.004	0.028	0.052
	(0.013)	(0.029)	(0.048)	(0.013)	(0.028)	(0.041)	(0.013)	(0.026)	(0.043)
Number of weight checks	0.1121**	0.087	0.2648*	0.1168**	0.049	0.056	-0.005	0.038	0.209
	(0.048)	(0.089)	(0.152)	(0.049)	(0.093)	(0.117)	(0.047)	(0.091)	(0.163)
Number Vitamin A supplements	0.067	-0.1574*	0.034	0.033	-0.1639*	0.020	0.033	0.007	0.014
	(0.042)	(0.092)	(0.100)	(0.053)	(0.099)	(0.100)	(0.051)	(0.079)	(0.124)
Percent malnourished	0.006	-0.0723**	-0.037	-0.002	-0.042	0.020	0.008	-0.030	-0.057
	(0.012)	(0.029)	(0.036)	(0.013)	(0.027)	(0.034)	(0.015)	(0.031)	(0.040)
Age 7–12 gross enrollment	-0.005	0.0275***	-0.001	-0.001	0.0287***	0.002	-0.004	-0.001	-0.003
	(0.004)	(0.008)	(0.015)	(0.004)	(0.009)	(0.013)	(0.004)	(0.009)	(0.017)
Age 13–15 gross enrollment	-0.004	-0.009	0.026	0.005	-0.022	-0.1014***	-0.009	0.013	0.1273***
	(0.017)	(0.037)	(0.040)	(0.020)	(0.036)	(0.037)	(0.018)	(0.032)	(0.035)
Age 7–12 gross attendance	-0.005	0.0322***	-0.016	-0.006	0.0305***	0.008	0.002	0.002	-0.024
	(0.005)	(0.010)	(0.018)	(0.005)	(0.010)	(0.014)	(0.005)	(0.011)	(0.020)
Age 13–15 gross attendance	-0.004	-0.007	0.018	0.006	-0.035	-0.1140***	-0.009	0.029	0.1316***
	(0.018)	(0.038)	(0.042)	(0.021)	(0.038)	(0.038)	(0.020)	(0.033)	(0.037)
Average standardized effect	0.010	0.052	0.0907*	0.024	-0.032	-0.057	-0.014	0.0831**	0.1478**
	(0.018)	(0.041)	(0.050)	(0.018)	(0.038)	(0.047)	(0.019)	(0.035)	(0.060)
Average standardized effect health	0.025	0.037	0.1316**	0.0365*	-0.071	-0.006	-0.012	0.1079***	0.1380**
	(0.020)	(0.044)	(0.053)	(0.021)	(0.046)	(0.054)	(0.020)	(0.038)	(0.070)
Average standardized effect educ.	-0.021	0.081	0.009	-0.002	0.048	-0.1584*	-0.019	0.034	0.1673**
	(0.031)	(0.066)	(0.074)	(0.036)	(0.064)	(0.081)	(0.034)	(0.057)	(0.076)

Notes: See Notes to Table 4. Each row reports a single regression with both treatment dummies interacted with NTT and Sulawesi dummies.

Table 6: Interactions with baseline level of service delivery

Indicator	Generasi Incentive Total Effect * Pre-Period Level	Generasi Non-Incentive Total Effect * Pre-Period Level	Generasi Incentive Additional Effect * Pre-Period Level	Incentive Additional Effect at 10th Percentile	Generasi Incentive Total Effect * Pre-Period Level	Generasi Non-Incentive Total Effect * Pre-Period Level	Generasi Incentive Additional Effect * Pre-Period Level	Incentive Additional Effect at 10th Percentile	Generasi Incentive Total Effect * Pre-Period Level	Generasi Non-Incentive Total Effect * Pre-Period Level	Generasi Incentive Additional Effect * Pre-Period Level	Incentive Additional Effect at 10th Percentile
	Wave II				Wave III				AVERAGE			
Main 12 indicators	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number prenatal visits	0.167	0.115	0.052	0.490	-0.045	-0.1456*	0.101	-0.095	0.065	-0.015	0.080	0.158
	(0.131)	(0.117)	(0.133)	(0.365)	(0.080)	(0.082)	(0.083)	(0.284)	(0.080)	(0.071)	(0.082)	(0.253)
Delivery by trained midwife	-0.088	0.042	-0.1296*	0.050	-0.056	0.049	-0.105	0.065	-0.074	0.040	-0.1143**	0.0588*
	(0.074)	(0.070)	(0.071)	(0.046)	(0.071)	(0.076)	(0.068)	(0.042)	(0.055)	(0.056)	(0.053)	(0.034)
Number of postnatal visits	-0.141	-0.039	-0.102	0.012	-0.2502*	-0.061	-0.189	0.221	-0.1904**	-0.043	-0.147	0.123
	(0.126)	(0.130)	(0.142)	(0.193)	(0.137)	(0.128)	(0.136)	(0.182)	(0.094)	(0.101)	(0.103)	(0.142)
Iron tablet sachets	-0.142	-0.206	0.064	0.044	0.009	0.160	-0.151	0.116	-0.082	-0.020	-0.061	0.086
	(0.126)	(0.130)	(0.152)	(0.111)	(0.124)	(0.143)	(0.153)	(0.097)	(0.093)	(0.093)	(0.111)	(0.074)
Percent of immunization	-0.1884**	-0.086	-0.102	0.041	0.016	0.074	-0.057	0.033	-0.102	-0.021	-0.080	0.0376*
	(0.085)	(0.078)	(0.087)	(0.029)	(0.079)	(0.073)	(0.066)	(0.025)	(0.062)	(0.057)	(0.056)	(0.020)
Number of weight checks	-0.071	-0.086	0.016	0.083	-0.065	0.000	-0.065	0.022	-0.069	-0.043	-0.025	0.045
	(0.098)	(0.099)	(0.106)	(0.108)	(0.110)	(0.115)	(0.129)	(0.120)	(0.069)	(0.086)	(0.094)	(0.093)
Number Vitamin A	-0.030	-0.024	-0.007	-0.008	-0.001	-0.044	0.043	0.060	-0.013	-0.037	0.024	0.026
Supplements	(0.128)	(0.115)	(0.154)	(0.096)	(0.115)	(0.130)	(0.133)	(0.085)	(0.085)	(0.093)	(0.106)	(0.065)
Percent malnourished	-0.2564**	-0.100	-0.156	-0.0481*	-0.2677**	-0.2400**	-0.028	0.006	-0.2591***	-0.1657**	-0.093	-0.020
	(0.129)	(0.113)	(0.138)	(0.027)	(0.132)	(0.116)	(0.128)	(0.027)	(0.095)	(0.078)	(0.100)	(0.021)
Age 7–12 gross enrollment	-0.042	-0.087	0.045	-0.007	-0.114	-0.1800**	0.066	-0.011	-0.074	-0.129	0.055	-0.009
	(0.090)	(0.106)	(0.127)	(0.012)	(0.094)	(0.081)	(0.098)	(0.010)	(0.066)	(0.080)	(0.091)	(0.009)
Age 13–15 gross enrollment	-0.063	-0.079	0.016	0.013	-0.006	-0.115	0.110	-0.021	-0.036	-0.100	0.065	-0.006
	(0.120)	(0.121)	(0.149)	(0.044)	(0.109)	(0.098)	(0.101)	(0.028)	(0.085)	(0.090)	(0.098)	(0.028)
Age 7–12 gross attendance	-0.051	-0.045	-0.006	0.000	-0.1064**	-0.1085**	0.002	-0.001	-0.0738**	-0.0721**	-0.002	-0.001
	(0.039)	(0.039)	(0.033)	(0.006)	(0.050)	(0.049)	(0.034)	(0.007)	(0.032)	(0.031)	(0.027)	(0.005)
Age 13–15 gross attendance	-0.052	-0.033	-0.019	0.031	-0.022	-0.110	0.087	-0.016	-0.037	-0.077	0.040	0.004
	(0.111)	(0.109)	(0.133)	(0.048)	(0.108)	(0.078)	(0.099)	(0.032)	(0.080)	(0.073)	(0.087)	(0.029)
Average standardized effect	-0.2140**	-0.157	-0.057	0.056	-0.1919**	-0.2225***	0.031	0.025	-0.2055***	-0.1957**	-0.010	0.037
	(0.096)	(0.112)	(0.133)	(0.042)	(0.091)	(0.085)	(0.088)	(0.033)	(0.065)	(0.079)	(0.087)	(0.029)
Average standardized effect	-0.1975***	-0.078	-0.1193*	0.0706*	-0.106	-0.020	-0.086	0.061	-0.1615***	-0.059	-0.1022**	0.0638**
health	(0.061)	(0.056)	(0.068)	(0.037)	(0.066)	(0.066)	(0.064)	(0.039)	(0.047)	(0.045)	(0.051)	(0.031)
Average standardized effect	-0.247	-0.314	0.067	0.026	-0.363	-0.6274***	0.264	-0.048	-0.2934*	-0.4685**	0.175	-0.017
educ.	(0.253)	(0.312)	(0.369)	(0.086)	(0.244)	(0.219)	(0.235)	(0.050)	(0.174)	(0.221)	(0.241)	(0.051)

Notes: See Notes to Table 4. Columns (1), (5), and (9) interact the incentive treatment dummy with the baseline subdistrict mean of the variable shown, and columns (2), (5), and (10) interact the non-incentive treatment dummy with the baseline subdistrict mean of the variable shown. Columns (3), (7), and (11) are the difference between the two previous columns. Columns (4), (8), and (12) show the previous column – the estimated additional impact of incentives – evaluated at the 10th percentile of the indicator at baseline.

Table 7: Spillovers on non-targeted indicators.

		Wave II			Wave III			AVERAGE		
Indicator		Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Average Treatment Effect	Non-Incentive Average Treatment Effect	Incentive Average Additional Effect
Health utilization										
Facility-based vs home	0.397	0.041	0.018	0.023	0.030	0.0431*	-0.013	0.0347*	0.0326*	0.002
Deliveries	[0.490]	(0.025)	(0.023)	(0.026)	(0.025)	(0.024)	(0.026)	(0.020)	(0.020)	(0.022)
Use of family planning	0.528	0.008	-0.007	0.016	-0.022	-0.0246*	0.003	-0.008	-0.017	0.008
	[0.499]	(0.016)	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)	(0.012)	(0.010)	(0.010)
Use of health services curative	.	-0.009	-0.001	-0.008	-0.007	-0.005	-0.002	-0.008	-0.003	-0.004
care	[.]	(0.018)	(0.019)	(0.018)	(0.021)	(0.020)	(0.020)	(0.014)	(0.014)	(0.014)
Any prenatal visits beyond 4	0.737	-0.008	0.002	-0.010	0.0340*	0.015	0.019	0.0272*	-0.008	0.0351**
	[0.440]	(0.011)	(0.011)	(0.011)	(0.019)	(0.020)	(0.018)	(0.015)	(0.015)	(0.015)
Any vitamin A beyond 2	0.043	-0.102	0.016	-0.118	0.016	-0.001	0.017	0.004	0.000	0.004
	[0.204]	(0.214)	(0.243)	(0.226)	(0.011)	(0.010)	(0.010)	(0.008)	(0.007)	(0.008)
Average standardized effect		0.019	-0.010	0.029	0.029	0.011	0.018	0.023	0.001	0.022
		(0.020)	(0.021)	(0.022)	(0.021)	(0.020)	(0.019)	(0.016)	(0.015)	(0.016)
Health quality										
Quality of prenatal care services	0.546	0.019	0.002	0.017	0.016	0.008	0.007	0.0171*	0.006	0.012
	[0.244]	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.010)	(0.010)	(0.009)
Quality of posyandu	.	0.0489*	0.0669***	-0.018	0.009	0.022	-0.013	0.0287*	0.0434**	-0.015
	[.]	(0.026)	(0.025)	(0.027)	(0.022)	(0.024)	(0.025)	(0.017)	(0.018)	(0.019)
Average standardized effect		0.0901**	0.0752*	0.015	0.041	0.040	0.001	0.0646**	0.0567**	0.008
		(0.039)	(0.039)	(0.040)	(0.036)	(0.038)	(0.036)	(0.029)	(0.029)	(0.028)
Maternal knowledge and practices										
Initiation of breastfeeding	0.575	0.023	0.008	0.015	0.031	0.022	0.009	0.027	0.016	0.012
	[0.494]	(0.023)	(0.022)	(0.023)	(0.022)	(0.020)	(0.020)	(0.017)	(0.016)	(0.017)
Exclusive breastfeeding	0.472	0.014	0.023	-0.008	-0.009	0.008	-0.017	0.002	0.015	-0.013
	[0.499]	(0.024)	(0.026)	(0.026)	(0.022)	(0.023)	(0.022)	(0.017)	(0.018)	(0.018)
Mother's knowledge	.	0.004	0.009	-0.005	0.013	0.0206***	-0.008	0.008	0.0151**	-0.007
	[.]	(0.009)	(0.009)	(0.010)	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)
Woman role in child decisions	0.773	-0.016	0.010	-0.026	0.006	-0.002	0.008	0.004	-0.003	0.008
dummy	[0.419]	(0.019)	(0.021)	(0.018)	(0.017)	(0.017)	(0.016)	(0.013)	(0.013)	(0.013)
Average standardized effect		0.026	0.024	0.002	0.033	0.043	-0.011	0.029	0.034	-0.005
		(0.029)	(0.028)	(0.030)	(0.029)	(0.027)	(0.026)	(0.022)	(0.022)	(0.021)
Family composition decisions										
Fertility rate	0.344	-0.001	-0.018	0.016	0.011	0.002	0.009	0.005	-0.006	0.012
	[0.475]	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)	(0.007)
Number migrate out village 12	.	-0.014	-0.008	-0.006	-1.642	1.261	-2.9026*	-1.642	1.261	-2.9026*
mons	[.]	(0.011)	(0.013)	(0.013)	(1.509)	(1.812)	(1.637)	(1.509)	(1.812)	(1.637)

Indicator		Wave II			Wave III			AVERAGE		
		Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Average Treatment Effect	Non-Incentive Average Treatment Effect	Incentive Average Additional Effect
Number HH indivs. Migrate out	0.069	0.022	-0.0342*	0.0558**	0.008	-0.009	0.017	-0.001	-0.009	0.008
kec 12 mons	[0.314]	(0.022)	(0.021)	(0.022)	(0.013)	(0.013)	(0.011)	(0.009)	(0.010)	(0.009)
Average standardized effect		0.014	-0.012	0.026	0.023	-0.007	0.029	0.025	-0.014	0.0381*
		(0.019)	(0.021)	(0.022)	(0.022)	(0.026)	(0.023)	(0.020)	(0.024)	(0.022)
Other education metrics										
Gross high school enrollment	.	-0.032	-0.022	-0.011	-0.0669**	-0.010	-0.0572*	-0.052	-0.014	-0.038
	[.]	(0.044)	(0.043)	(0.045)	(0.034)	(0.033)	(0.034)	(0.032)	(0.031)	(0.034)
Dropout rates	0.026	0.007	0.007	-0.001	0.000	-0.002	0.002	0.003	0.002	0.001
	[0.158]	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
SD to SMP transition	0.885	-0.009	0.000	-0.008	0.010	0.016	-0.006	0.001	0.009	-0.007
	[0.319]	(0.014)	(0.010)	(0.013)	(0.020)	(0.018)	(0.015)	(0.012)	(0.011)	(0.011)
Number of hours attended school	17.112	-0.6103*	-0.8002*	0.190	-0.147	0.123	-0.270	-0.369	-0.279	-0.090
	[13.674]	(0.343)	(0.415)	(0.441)	(0.475)	(0.432)	(0.434)	(0.329)	(0.330)	(0.345)
Numbers attend Paket A					0.000	-0.0008*	0.001	0.000	-0.0008*	0.001
					(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Numbers attend Paket B					0.000	0.000	0.000	0.000	0.000	0.000
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Numbers attend Paket C					0.001	-0.0009*	0.0014**	0.001	-0.0009*	0.0014**
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average standardized effect		-0.070	-0.051	-0.019	-0.013	0.006	-0.019	-0.022	-0.011	-0.012
		(0.049)	(0.046)	(0.049)	(0.021)	(0.020)	(0.018)	(0.017)	(0.018)	(0.018)
Distance to school										
Distance to SMP (km)	11.439	0.0501**	0.024	0.026	0.076	0.005	0.071	0.001	0.003	-0.003
	[20.544]	(0.024)	(0.022)	(0.031)	(0.203)	(0.198)	(0.196)	(0.165)	(0.169)	(0.166)
Time spent oneway to SMP (hr)	0.397	144.559	25.835	118.724	-0.008	-0.001	-0.007	0.017	0.011	0.006
	[0.350]	(140.376)	(116.831)	(139.557)	(0.020)	(0.020)	(0.021)	(0.018)	(0.016)	(0.020)
Transportation cost oneway to	777.527	0.003	-0.005	0.008	-35.299	-180.402	145.103	45.869	-92.027	137.897
SMP	[1421.307]	(0.019)	(0.017)	(0.018)	(153.749)	(150.412)	(122.304)	(107.377)	(104.765)	(101.851)
Average standardized effect		-0.077	-0.034	-0.043	0.004	0.022	-0.018	-0.025	0.002	-0.027
		(0.058)	(0.050)	(0.060)	(0.042)	(0.041)	(0.042)	(0.036)	(0.035)	(0.039)
Child labor (note these are bad so avg std effect * -1)										
Age 7-15 hours wage	0.431	-0.003	0.3647**	-0.3672***	0.023	-0.012	0.035	0.018	0.1459**	-0.1275**
work	[3.869]	(0.078)	(0.143)	(0.125)	(0.076)	(0.072)	(0.066)	(0.055)	(0.073)	(0.064)
Age 7-15 hours household work	3.915	0.4178*	0.9241***	-0.5063*	-0.161	-0.150	-0.012	0.114	0.3196*	-0.206
	[6.761]	(0.227)	(0.291)	(0.288)	(0.212)	(0.189)	(0.196)	(0.155)	(0.173)	(0.173)
Age 7-15 wage work dummy	0.031	-0.001	0.0126*	-0.0132**	0.003	-0.002	0.005	0.001	0.004	-0.003
	[0.174]	(0.005)	(0.007)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)

		Wave II			Wave III			AVERAGE		
Indicator		Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Average Treatment Effect	Non-Incentive Average Treatment Effect	Incentive Average Additional Effect
Age 7-15 household work dummy	0.728 [0.445]	0.013 (0.016)	0.002 (0.017)	0.011 (0.018)	-0.022 (0.020)	0.006 (0.019)	-0.029 (0.022)	-0.007 (0.014)	0.006 (0.014)	-0.012 (0.016)
Average standardized effect		-0.025 (0.022)	-0.1074*** (0.038)	0.0825** (0.034)	0.012 (0.025)	0.007 (0.020)	0.005 (0.022)	-0.006 (0.018)	-0.0391* (0.021)	0.0335* (0.020)
Average standardized effect		-0.003 (0.016)	-0.017 (0.017)	0.013 (0.019)	0.012 (0.025)	0.007 (0.020)	0.005 (0.022)	0.013 (0.010)	0.004 (0.011)	0.008 (0.011)
Average standardized effect health		0.0373** (0.016)	0.019 (0.016)	0.018 (0.017)	0.012 (0.025)	0.007 (0.020)	0.005 (0.022)	0.0354*** (0.013)	0.020 (0.013)	0.016 (0.013)
Average standardized effect educ.		-0.0574** (0.029)	-0.0643** (0.030)	0.007 (0.032)	0.012 (0.025)	0.007 (0.020)	0.005 (0.022)	-0.018 (0.016)	-0.016 (0.016)	-0.002 (0.017)

Notes: See Notes to Table 4.

Table 8: Recordkeeping

Indicator	Baseline mean	Wave II: Incentive Treatment Effect	Wave II: Non-Incentive Treatment Effect	Wave II: Incentive Additional Effect	Wave III: Incentive Treatment Effect	Wave III: Non-Incentive Treatment Effect	Wave III: Incentive Additional Effect	Average: Incentive Average Treatment Effect	Average: Non-Incentive Average Treatment Effect	Average: Incentive Average Additional Effect
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: BCG Scar										
False "yes" in recorded BCG vaccine	0.100	0.0304**	0.006	0.025	-0.002	0.002	-0.003	0.009	0.001	0.008
	[0.2995]	(0.015)	(0.014)	(0.015)	(0.013)	(0.014)	(0.014)	(0.010)	(0.011)	(0.011)
False "yes" in declared BCG vaccine	0.122	0.0297*	0.019	0.010	0.007	0.000	0.008	0.014	0.005	0.009
	[0.3274]	(0.015)	(0.015)	(0.016)	(0.014)	(0.015)	(0.014)	(0.011)	(0.011)	(0.011)
False "no" in recorded BCG vaccine	0.127	0.019	0.022	-0.003	-0.002	-0.011	0.009	0.011	0.005	0.006
	[0.3329]	(0.020)	(0.019)	(0.022)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)
False "no" in declared BCG vaccine	0.011	-0.001	0.000	-0.001	-0.001	-0.002	0.001	-0.002	-0.002	0.000
	[0.1056]	(0.008)	(0.008)	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Dummy for having a BCG scar	0.746	-0.029	-0.034	0.005	-0.010	0.007	-0.017	-0.018	-0.007	-0.010
	[0.4358]	(0.020)	(0.023)	(0.022)	(0.016)	(0.017)	(0.015)	(0.013)	(0.014)	(0.013)
Dummy for recorded BCG vaccine	0.604	0.004	-0.015	0.019	0.001	0.0388*	-0.0382*	-0.001	0.017	-0.018
	[0.4893]	(0.026)	(0.026)	(0.025)	(0.022)	(0.021)	(0.020)	(0.018)	(0.017)	(0.018)
Dummy for declared BCG vaccine	0.845	0.004	-0.015	0.019	0.003	0.006	-0.003	0.004	-0.002	0.006
	[0.3620]	(0.018)	(0.018)	(0.018)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
Children with no record card	0.195	-0.0481**	-0.0349*	-0.013	-0.020	-0.0515***	0.0314*	-0.0289*	-0.0423***	0.013
	[0.3963]	(0.021)	(0.021)	(0.019)	(0.019)	(0.018)	(0.017)	(0.015)	(0.015)	(0.014)
Panel B: Attendance										
Attend. Rate – difference between recorded and observed	7.299	-2.088	-2.6623*	0.574	0.496	2.158	-1.663	-0.589	0.245	-0.834
		(1.660)	(1.500)	(1.686)	(2.018)	(2.119)	(1.971)	(1.324)	(1.266)	(1.295)
Attend. rate observed	88.366	1.348	2.9302**	-1.582	-0.795	-1.975	1.180	0.070	-0.036	0.106
		(1.577)	(1.466)	(1.617)	(1.885)	(2.024)	(1.900)	(1.254)	(1.233)	(1.266)
Attend. rate recorded	95.726	-0.7465**	0.164	-0.9106**	-0.253	0.160	-0.414	-0.472	0.168	-0.6395*
		(0.368)	(0.399)	(0.459)	(0.441)	(0.435)	(0.441)	(0.306)	(0.308)	(0.333)

Notes: See Notes to Table 4.

Table 9: Change in budget allocations

Indicator	Wave II Incentive Mean	Wave II Non-Incentive Mean	Wave II Incentive Additional Effect	Wave III Incentive Mean	Wave III Non-Incentive Mean	Wave III Incentive Additional Effect	AVERAGE Incentive Mean	AVERAGE Non-Incentive Mean	AVERAGE Incentive Average Additional Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Health vs. education									
All health expenditures	0.4696	0.4320	0.0308*	0.4904	0.4672	0.0304**	0.4812	0.4521	0.0306***
			(0.016)			(0.012)			(0.011)
Health durables	0.0985	0.0855	0.014	0.1262	0.1086	0.019	0.1140	0.0983	0.0168*
			(0.012)			(0.012)			(0.010)
Health benefiting providers	0.0166	0.0141	0.004	0.0220	0.0224	0.002	0.0196	0.0192	0.003
			(0.005)			(0.003)			(0.004)
Panel B: Transfers									
All transfers	0.7312	0.7563	-0.036	0.7281	0.7448	-0.016	0.7295	0.7502	-0.023
			(0.025)			(0.021)			(0.019)
Education supplies	0.2359	0.2745	-0.0463*	0.2360	0.2721	-0.0307*	0.2360	0.2723	-0.0379**
			(0.024)			(0.018)			(0.017)
Supplementary feeding	0.2166	0.1771	0.018	0.2122	0.2130	0.004	0.2141	0.1977	0.010
			(0.014)			(0.012)			(0.010)
Subsidies	0.2787	0.3048	-0.007	0.2799	0.2597	0.012	0.2793	0.2802	0.005
			(0.024)			(0.019)			(0.018)

Notes: See Notes to Table 4. A unit of observation is a village. Since budget data is (by definition) only available for treatment areas, columns (3) , (6), and (9) regress the variable shown on a variable for being an incentive subdistrict. Robust standard errors in parentheses, adjusted for clustering at the subdistrict level.

Table 10: Direct benefits, incentivized vs. non-incentivized

Indicator	Control Mean	Wave II: Incentive Treatment Effect	Wave II: Non-Incentive Treatment Effect	Wave II: Incentive Additional Effect	Wave III: Incentive Treatment Effect	Wave III: Non-Incentive Treatment Effect	Wave III: Incentive Additional Effect	AVERAGE: Incentive Average Treatment Effect	AVERAGE: Non-Incentive Average Treatment Effect	AVERAGE: Incentive Average Additional Effect
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Received scholarship	0.025	0.0162**	0.008	0.009	0.0208**	0.009	0.012	0.0187***	0.008	0.0108*
	[0.0048]	(0.007)	(0.006)	(0.008)	(0.009)	(0.007)	(0.009)	(0.005)	(0.005)	(0.006)
Received uniform	0.013	0.1095***	0.0828***	0.027	0.0824***	0.0723***	0.010	0.0948***	0.0777***	0.017
	[0.0036]	(0.019)	(0.012)	(0.018)	(0.013)	(0.010)	(0.015)	(0.013)	(0.008)	(0.012)
Received other school	0.008	0.0634***	0.0535***	0.010	0.0701***	0.0534***	0.017	0.0670***	0.0533***	0.014
supplies	[0.0027]	(0.012)	(0.009)	(0.012)	(0.012)	(0.010)	(0.015)	(0.010)	(0.007)	(0.011)
Received transport	0.000	0.0143***	0.0049*	0.009	0.0078***	0.0050***	0.003	0.0108***	0.0051***	0.0056*
subsidy	[0.0000]	(0.005)	(0.003)	(0.006)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Received other school	0.000	0.000	0.000	0.001	0.0072**	0.0063*	0.001	0.0039**	0.0033*	0.001
support	[0.0000]	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Received supp. feeding at	0.000	0.005	0.0041**	0.001	0.006	0.003	0.003	0.005	0.004	0.002
school	[0.0000]	(0.003)	(0.002)	(0.004)	(0.006)	(0.005)	(0.007)	(0.004)	(0.003)	(0.004)
Received supp. feeding at	0.469	0.1533***	0.1563***	-0.003	0.1745***	0.2044***	-0.030	0.1647***	0.1843***	-0.020
posyandu	[0.0171]	(0.028)	(0.027)	(0.028)	(0.025)	(0.022)	(0.023)	(0.022)	(0.019)	(0.019)
Received intensive supp.	0.027	0.008	0.0252**	-0.018	0.0242**	0.0191**	0.005	0.0173***	0.0212***	-0.004
feeding at school	[0.0055]	(0.007)	(0.011)	(0.011)	(0.010)	(0.009)	(0.010)	(0.007)	(0.007)	(0.007)
Received health subsidy	0.005	0.0343***	0.0270***	0.007	0.0273***	0.0364***	-0.009	0.0304***	0.0323***	-0.002
for pre/postnatal care	[0.0023]	(0.008)	(0.007)	(0.009)	(0.006)	(0.007)	(0.009)	(0.006)	(0.006)	(0.007)
Received health subsidy	0.039	0.1010***	0.1273***	-0.026	0.0974***	0.1249***	-0.028	0.0991***	0.1260***	-0.027
for childbirth	[0.0078]	(0.017)	(0.017)	(0.019)	(0.016)	(0.020)	(0.023)	(0.012)	(0.015)	(0.016)
Average standardized		0.3394***	0.2995***	0.040	0.3076***	0.2950***	0.013	0.3526***	0.3140***	0.039
effect		(0.041)	(0.030)	(0.040)	(0.031)	(0.028)	(0.039)	(0.032)	(0.026)	(0.035)
Average standardized		0.2847***	0.3122***	-0.028	0.2657***	0.3136***	-0.048	0.3179***	0.3620***	-0.044
effect health		(0.037)	(0.031)	(0.039)	(0.031)	(0.035)	(0.042)	(0.039)	(0.040)	(0.047)
Average standardized		0.3940***	0.2867***	0.1073*	0.3495***	0.2764***	0.073	0.3734***	0.2852***	0.0883*
effect educ.		(0.063)	(0.041)	(0.060)	(0.049)	(0.041)	(0.059)	(0.045)	(0.032)	(0.046)

Note: See Notes to Table 4. Note that instead of showing a baseline mean, we show the wave II control group mean because there is no data available for these categories in Wave I. These regressions also therefore do not control for baseline values.

Table 11: Worker Behavior

Indicator	Baseline Mean	Wave II: Incentive Treatment Effect	Wave II: Non-Incentive Treatment Effect	Wave II: Incentive Additional Effect	Wave III: Incentive Treatment Effect	Wave III: Non-Incentive Treatment Effect	Wave III: Incentive Additional Effect	AVERAGE: Incentive Average Treatment Effect	AVERAGE: Non-Incentive Average Treatment Effect	AVERAGE: Incentive Average Additional Effect
Midwives:										
Hours spent in outreach over past 3	3.165	0.7961*	-0.074	0.8700**	0.076	0.038	0.038	0.391	0.007	0.383
days	[4.4875]	(0.410)	(0.337)	(0.425)	(0.389)	(0.419)	(0.400)	(0.299)	(0.305)	(0.327)
Hours spent providing public services	13.548	0.536	-1.1020*	1.6380**	0.675	0.417	0.257	0.579	-0.248	0.8272*
over past 3 days	[10.0559]	(0.608)	(0.594)	(0.721)	(0.619)	(0.567)	(0.585)	(0.460)	(0.419)	(0.487)
Hours spent providing private services	10.805	0.212	-0.469	0.681	0.894	0.591	0.304	0.570	0.112	0.458
over past 3 days	[12.5048]	(0.832)	(0.826)	(0.886)	(0.674)	(0.669)	(0.644)	(0.525)	(0.524)	(0.524)
Total hours spent working over past 3	27.518	1.477	-1.7182*	3.1956***	1.6276*	0.936	0.692	1.5004**	-0.224	1.7246**
days	[15.7132]	(1.047)	(1.039)	(1.154)	(0.951)	(0.932)	(0.884)	(0.712)	(0.728)	(0.723)
Number of posyandus attended in past	4.166	0.189	0.059	0.130	-0.162	0.053	-0.215	-0.009	0.064	-0.073
month	[3.3213]	(0.332)	(0.227)	(0.348)	(0.247)	(0.268)	(0.324)	(0.241)	(0.195)	(0.294)
Number of hours midwife per										
posyandu	3.039	0.137	0.181	-0.044	0.110	-0.083	0.192	0.127	0.032	0.095
Teachers:	[1.6932]	(0.130)	(0.120)	(0.127)	(0.152)	(0.133)	(0.153)	(0.111)	(0.095)	(0.111)
Percent present at time of interview	.	0.006	0.016	-0.010	0.000	0.008	-0.009	0.006	0.016	-0.010
(primary)	[.]	(0.016)	(0.015)	(0.017)	(0.011)	(0.011)	(0.012)	(0.016)	(0.015)	(0.017)
Percent present at time of interview	.	0.001	-0.010	0.011	-0.008	-0.015	0.007	-0.004	-0.013	0.009
(junior secondary)	[.]	(0.015)	(0.014)	(0.014)	(0.012)	(0.012)	(0.013)	(0.010)	(0.010)	(0.010)
Percent observed teaching (primary)	.	-0.006	-0.050	0.044	-0.003	-0.012	0.009	-0.005	-0.028	0.023
	[.]	(0.038)	(0.042)	(0.042)	(0.040)	(0.041)	(0.038)	(0.028)	(0.029)	(0.028)
Percent observed teaching (j. sec.)	.	-0.069	-0.052	-0.018	0.039	0.024	0.015	-0.010	-0.011	0.002
Puskesmas:	[.]	(0.044)	(0.047)	(0.049)	(0.049)	(0.048)	(0.044)	(0.033)	(0.033)	(0.032)
Minutes wait at recent health visits	25.201	0.778	6.035	-5.257	2.409	1.281	1.128	1.696	3.042	-1.345
	[23.7360]	(3.637)	(4.685)	(3.953)	(4.269)	(4.224)	(4.400)	(3.033)	(3.302)	(3.320)
Percent of providers present at time of	.	0.0714**	0.1090***	-0.038	-0.009	-0.0757**	0.0667**	0.030	0.006	0.024
observation	[.]	(0.036)	(0.039)	(0.035)	(0.029)	(0.030)	(0.030)	(0.022)	(0.023)	(0.024)
Average standardized effect		0.045	-0.040	0.0846***	0.043	0.021	0.021	0.0409*	-0.005	0.0463*
		(0.029)	(0.028)	(0.032)	(0.028)	(0.028)	(0.030)	(0.022)	(0.020)	(0.024)
Average standardized effect health		0.0892**	-0.036	0.1250***	0.056	0.030	0.026	0.0665**	0.000	0.0662**
		(0.044)	(0.038)	(0.048)	(0.040)	(0.039)	(0.040)	(0.031)	(0.028)	(0.034)
Average standardized effect educ.		-0.022	-0.046	0.024	0.023	0.009	0.014	0.002	-0.014	0.016
		(0.043)	(0.044)	(0.047)	(0.041)	(0.042)	(0.042)	(0.030)	(0.029)	(0.030)

Note: See Notes to Table 4.

Table 12: Community effort

Indicator	Baseline Mean	Wave II: Incentive Treatment Effect	Wave II: Non-Incentive Treatment Effect	Wave II: Incentive Additional Effect	Wave III: Incentive Treatment Effect	Wave III: Non-Incentive Treatment Effect	Wave III: Incentive Additional Effect	AVERAGE: Incentive Average Treatment Effect	AVERAGE: Non-Incentive Average Treatment Effect	AVERAGE: Incentive Average Additional Effect
Community effort at direct service provision:										
Number of posyandus in village	4.5191	-0.092	0.004	-0.096	0.128	0.196	-0.068	0.027	0.107	-0.080
	[3.5043]	(0.124)	(0.147)	(0.126)	(0.178)	(0.176)	(0.148)	(0.140)	(0.151)	(0.120)
Number of posyandu meetings in	.	-0.003	0.082	-0.084	-0.112	-0.063	-0.049	-0.061	0.002	-0.063
past year at selected posyandu	[.]	(0.102)	(0.111)	(0.102)	(0.112)	(0.091)	(0.100)	(0.079)	(0.076)	(0.078)
Number of cadres at posyandu	.	0.174	0.197	-0.023	0.2890**	0.3577**	-0.069	0.2349**	0.2854**	-0.051
	[.]	(0.113)	(0.153)	(0.138)	(0.139)	(0.171)	(0.165)	(0.105)	(0.139)	(0.133)
Community effort at outreach										
Number of sweepings at selected	.	-0.296	0.042	-0.338	-0.127	-0.6155*	0.4888*	-0.186	-0.337	0.150
posyandu in last year	[.]	(0.394)	(0.377)	(0.389)	(0.342)	(0.346)	(0.295)	(0.266)	(0.257)	(0.257)
Number of primary school comm..	.	0.066	-0.070	0.136	0.002	-0.125	0.126	0.031	-0.099	0.130
meetings with parents in past year	[.]	(0.133)	(0.133)	(0.121)	(0.181)	(0.182)	(0.137)	(0.117)	(0.119)	(0.093)
Number of junior sec. school	2.3093	-0.121	0.032	-0.153	0.213	0.210	0.003	0.066	0.125	-0.060
committee meetings w parents	[1.9728]	(0.113)	(0.118)	(0.126)	(0.147)	(0.223)	(0.207)	(0.103)	(0.147)	(0.140)
Community effort at monitoring										
Number of primary school	.	0.7613*	-0.503	1.2638***	-0.003	0.195	-0.198	0.317	-0.085	0.401
committee members	[.]	(0.392)	(0.410)	(0.478)	(0.334)	(0.402)	(0.344)	(0.287)	(0.314)	(0.297)
Number of junior sec school	8.2592	-0.845	-1.421	0.577	0.199	0.231	-0.032	-0.296	-0.511	0.215
committee members	[4.7625]	(0.993)	(0.934)	(0.539)	(0.332)	(0.332)	(0.291)	(0.498)	(0.475)	(0.297)
Number of prim. school committee	.	-0.124	-0.367	0.243	-0.121	-0.096	-0.025	-0.129	-0.213	0.084
meetings with teachers in past year	[.]	(0.358)	(0.357)	(0.354)	(0.316)	(0.319)	(0.268)	(0.255)	(0.252)	(0.211)
Number of j. sec. school committee	4.4761	0.477	0.132	0.345	0.530	0.5755*	-0.045	0.4957*	0.381	0.115
meetings with teachers in year	[5.4650]	(0.424)	(0.394)	(0.455)	(0.342)	(0.346)	(0.364)	(0.262)	(0.258)	(0.269)
Average standardized effect		0.014	-0.009	0.023	0.0431*	0.048	-0.004	0.026	0.017	0.010
		(0.022)	(0.025)	(0.023)	(0.025)	(0.031)	(0.029)	(0.018)	(0.022)	(0.019)

Note: See Notes to Table 4.

Table 13: Cost-effectiveness Calculation

	Generasi with Incentives	Generasi without Incentives	Additional effect of incentives	Conditional Cash Transfer (PKH) (no spillover)	Conditional Cash Transfer (PKH) (w. spillover)	Generasi with Incentives	Generasi without Incentives	Additional effect of incentives	Conditional Cash Transfer (PKH) (no spillover)	Conditional Cash Transfer (PKH) (w. spillover)
Panel A: Social Cost Effectiveness	Entire program									
Transfers	0.00	0.00	0.00	0.00	0.00					
Non-transfers	3.91	3.68	0.23	0.00	0.00					
Facilitation	2.54	2.54	0.00	18.40	18.40					
Marginal cost of public funds	5.07	5.07	0.00	32.04	32.04					
Total costs (millions USD)	11.51	11.28	0.23	50.44	50.44					
Millions of points	1.42	1.04	0.373	2.24	4.46					
Dollars per point	8.13	10.81	0.62	22.43	11.30					
Panel B: Social Cost Effectiveness by Area	Health					Education				
Transfers	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Non-transfers	2.24	1.97	0.28	0.00	0.00	1.63	1.68	-0.05	0.00	0.00
Facilitation	1.27	1.27	0.00	21.72	21.72	1.27	1.27	0.00	21.72	21.72
Marginal cost of public funds	2.87	2.77	0.10	16.02	16.02	2.95	3.05	-0.10	16.02	16.02
Total costs (millions USD)	6.39	6.01	0.38	37.74	37.74	5.85	6.00	-0.15	37.74	37.74
Millions of points	0.96	0.67	0.291	2.25	4.47	0.46	0.38	0.081	0.00	0.00
Dollars per point	6.66	9.01	1.30	16.78	8.45	12.78	15.93	N/A	N/A	N/A
Panel C: Dollars Spent Cost Effectiveness	Entire program, not including transfers					Entire program, including transfers				
Transfers	0.00	0.00	0.00	0.00	0.00	10.46	10.69	-0.23	88.40	88.40
Non-transfers	3.91	3.68	0.23	0.00	0.00	3.91	3.68	0.23	0.00	0.00
Facilitation	2.54	2.54	0.00	18.40	18.40	2.54	2.54	0.00	18.40	18.40
Marginal cost of public funds	0.00	0.00	0.00			0.00	0.00	0.00		
Total costs (millions USD)	6.44	6.21	0.23	106.80	106.80	16.90	16.90	0.373	106.80	106.80
Millions of points	1.42	1.04	0.373	2.458	4.466	1.42	1.04	0.00	2.458	4.466
Dollars per point	4.55	5.95	0.62	7.49	4.12	11.93	16.20	0.00	43.45	23.92

Notes: Note that the costs and points for Generasi have been divided by 2, so that in this calculation exactly half the benefits and costs have been allocated to the program with and without incentives. The estimated points are therefore 50% of the estimated numbers in Table 4 above. PKH calculations are authors calculations based on the coefficients given in Alatas et. al (2010), as well as authors' calculations of the average number of beneficiaries of different age ranges per PKH household based on the PKH wave 3 survey. For health and education, we allocate the facilitation costs and PKH transfers 50-50 between health and education, and allocate actual Generasi expenditures based on the actual distribution of expenditures between health and education in the MIS data.

Appendix Tables

Appendix Table 1: Robustness of main results to alternative specifications

Wave II

Indicator	(1) Baseline mean	(2) Control mean	(3) Main specification	(4) Baseline controls for all 12 indicators	(5) Baseline controls for kecamatan averages only (no individual panel)	(6) No controls	(7) First differences	(8) Kecamatan level regression, with baseline control	(9) Full Intent-to-treat on 300 kecamatan, controlling for kecamatan avg
Number prenatal visits	7.447	7.464	0.6129***	0.5661***	0.5885***	0.4963**	0.8315***	0.5883**	0.5219**
	[4.2935]	[4.1639]	(0.220)	(0.216)	(0.218)	(0.236)	(0.273)	(0.241)	(0.209)
Delivery by trained midwife	0.670	0.755	-0.005	0.000	-0.001	0.005	-0.011	0.010	0.007
	[0.4705]	[0.4303]	(0.025)	(0.024)	(0.026)	(0.030)	(0.032)	(0.027)	(0.026)
Number of postnatal visits	1.720	1.737	-0.104	-0.146	-0.090	-0.106	-0.059	-0.105	-0.073
	[2.4477]	[2.4079]	(0.140)	(0.136)	(0.142)	(0.144)	(0.185)	(0.159)	(0.137)
Iron tablet sachets	1.588	1.977	0.078	0.071	0.077	0.070	0.135	0.074	0.056
	[1.2554]	[1.4426]	(0.081)	(0.081)	(0.081)	(0.081)	(0.102)	(0.085)	(0.077)
Percent of immunization	0.653	0.693	0.015	0.015	0.015	0.013	0.029	0.012	0.008
	[0.3664]	[0.3441]	(0.018)	(0.018)	(0.018)	(0.020)	(0.027)	(0.022)	(0.018)
Number of weight checks	2.126	2.192	0.0958*	0.1010*	0.0970*	0.085	0.1310*	0.1125*	0.1113**
	[1.1895]	[1.1718]	(0.054)	(0.056)	(0.054)	(0.058)	(0.072)	(0.061)	(0.053)
Number Vitamin A supplements	1.529	1.560	-0.013	-0.013	-0.004	0.000	-0.077	-0.019	0.016
	[1.1370]	[1.0089]	(0.058)	(0.057)	(0.059)	(0.062)	(0.090)	(0.066)	(0.060)
Percent malnourished	0.168	0.199	-0.0265*	-0.023	-0.024	-0.023	-0.0396*	-0.018	-0.021
	[0.3739]	[0.3995]	(0.016)	(0.016)	(0.016)	(0.015)	(0.023)	(0.017)	(0.015)
Age 7–12 gross enrollment	0.948	0.982	-0.004	-0.004	-0.004	-0.002	-0.0184*	-0.004	-0.006
	[0.2221]	[0.1334]	(0.006)	(0.005)	(0.006)	(0.006)	(0.011)	(0.006)	(0.005)
Age 13–15 gross enrollment	0.822	0.906	0.016	0.013	0.012	0.019	-0.003	-0.003	0.023
	[0.3827]	[0.2928]	(0.024)	(0.022)	(0.025)	(0.025)	(0.035)	(0.026)	(0.024)
Age 7–12 gross attendance	0.904	0.956	-0.001	-0.001	-0.001	0.000	-0.012	-0.001	-0.006
	[0.2773]	[0.1568]	(0.006)	(0.006)	(0.006)	(0.006)	(0.026)	(0.006)	(0.006)
Age 13–15 gross attendance	0.768	0.884	0.025	0.025	0.012	0.021	-0.025	-0.003	0.023
	[0.4125]	[0.3022]	(0.025)	(0.023)	(0.026)	(0.026)	(0.040)	(0.028)	(0.025)
Average standardized effect			0.036	0.033	0.032	0.035	0.017	0.049	0.034
			(0.024)	(0.022)	(0.024)	(0.027)	(0.029)	(0.060)	(0.023)
Average standardized effect health			0.0406*	0.0374*	0.0418*	0.038	0.0545**	0.0946*	0.0418*
			(0.024)	(0.023)	(0.024)	(0.029)	(0.027)	(0.057)	(0.023)
Average standardized effect educ.			0.027	0.024	0.013	0.031	-0.058	-0.044	0.019
			(0.045)	(0.041)	(0.047)	(0.049)	(0.067)	(0.119)	(0.045)

Note: See Notes to Table 4. Reported coefficients are the coefficient on the additional effect of the incentives.

Wave III

Indicator	(1)x Baseline mean	(2)x Control mean	(3)x Main specification	(4)x Baseline controls for all 12 indicators	(5)x Baseline controls for kecamatan averages only (no individual panel)	(6)x No controls	(7)x First differences	(8) Kecamatan level regression, with baseline control	(9) Full Intent-to-treat on 300 kecamatan, controlling for kecamatan avg
Number prenatal visits	7.447	7.639	0.181	0.187	0.186	0.134	0.368	0.231	0.098
	[4.2935]	[4.2297]	(0.173)	(0.174)	(0.173)	(0.173)	(0.257)	(0.187)	(0.165)
Delivery by trained midwife	0.670	0.780	0.019	0.018	0.021	0.021	0.020	0.015	0.018
	[0.4705]	[0.4144]	(0.021)	(0.021)	(0.021)	(0.022)	(0.031)	(0.023)	(0.021)
Number of postnatal visits	1.720	1.634	-0.003	0.032	-0.001	-0.017	0.057	-0.048	-0.017
	[2.4477]	[2.4597]	(0.129)	(0.129)	(0.129)	(0.129)	(0.163)	(0.137)	(0.122)
Iron tablet sachets	1.588	1.741	0.031	0.026	0.032	0.029	0.083	0.026	0.002
	[1.2554]	[1.2748]	(0.063)	(0.064)	(0.063)	(0.064)	(0.083)	(0.069)	(0.060)
Percent of immunization	0.653	0.756	0.017	0.017	0.017	0.012	0.025	0.008	0.013
	[0.3664]	[0.2858]	(0.014)	(0.014)	(0.014)	(0.015)	(0.020)	(0.016)	(0.013)
Number of weight checks	2.126	2.262	-0.024	-0.017	-0.024	-0.052	0.058	-0.046	-0.021
	[1.1895]	[1.1204]	(0.051)	(0.049)	(0.051)	(0.052)	(0.068)	(0.054)	(0.048)
Number Vitamin A supplements	1.529	1.454	0.083	0.1059**	0.083	0.055	-0.012	0.005	0.070
	[1.1370]	[0.9520]	(0.053)	(0.051)	(0.053)	(0.057)	(0.081)	(0.067)	(0.052)
Percent malnourished	0.168	0.228	0.009	0.007	0.009	0.009	0.001	0.009	0.006
	[0.3739]	[0.4199]	(0.016)	(0.017)	(0.016)	(0.016)	(0.023)	(0.017)	(0.015)
Age 7–12 gross enrollment	0.948	0.985	-0.006	-0.004	-0.005	-0.005	-0.012	-0.004	-0.005
	[0.2221]	[0.1207]	(0.005)	(0.004)	(0.005)	(0.005)	(0.009)	(0.005)	(0.004)
Age 13–15 gross enrollment	0.822	0.874	0.007	0.011	0.011	0.011	0.034	0.011	0.019
	[0.3827]	[0.3327]	(0.014)	(0.014)	(0.014)	(0.015)	(0.026)	(0.017)	(0.014)
Age 7–12 gross attendance	0.904	0.960	-0.001	0.000	-0.001	-0.001	0.006	0.002	0.000
	[0.2773]	[0.1463]	(0.006)	(0.006)	(0.006)	(0.006)	(0.021)	(0.006)	(0.006)
Age 13–15 gross attendance	0.768	0.860	0.010	0.019	0.012	0.017	0.002	0.015	0.017
	[0.4125]	[0.3364]	(0.015)	(0.015)	(0.015)	(0.016)	(0.033)	(0.019)	(0.015)
Average standardized effect			0.018	0.027	0.021	0.015	0.032	0.020	0.018
			(0.019)	(0.019)	(0.019)	(0.020)	(0.027)	(0.052)	(0.018)
Average standardized effect health			0.026	0.032	0.027	0.016	0.041	0.013	0.018
			(0.022)	(0.022)	(0.022)	(0.023)	(0.028)	(0.052)	(0.021)
Average standardized effect educ.			0.003	0.016	0.008	0.013	0.016	0.035	0.020
			(0.027)	(0.027)	(0.027)	(0.027)	(0.051)	(0.093)	(0.027)

Pooled

Indicator	(1)x Baseline mean	(2)x Control mean	(3)x Main specification	(4)x Baseline controls for all 12 indicators	(5)x Baseline controls for kecamatan averages only (no individual panel)	(6)x No controls	(7)x First differences	(8) Kecamatan level regression, with baseline control	(9) Full Intent-to-treat on 300 kecamatan, controlling for kecamatan avg
Number prenatal visits	7.447	7.551	0.3706**	0.3547**	0.3622**	0.2925*	0.5711**	0.3782**	0.2811**
	[4.2935]	[4.1963]	(0.147)	(0.146)	(0.147)	(0.155)	(0.225)	(0.148)	(0.143)
Delivery by trained midwife	0.670	0.768	0.009	0.011	0.012	0.015	0.007	0.013	0.014
	[0.4705]	[0.4223]	(0.018)	(0.018)	(0.019)	(0.021)	(0.027)	(0.017)	(0.018)
Number of postnatal visits	1.720	1.685	-0.048	-0.047	-0.041	-0.056	0.003	-0.072	-0.043
	[2.4477]	[2.4339]	(0.101)	(0.099)	(0.102)	(0.105)	(0.144)	(0.104)	(0.096)
Iron tablet sachets	1.588	1.860	0.052	0.047	0.052	0.047	0.106	0.046	0.026
	[1.2554]	[1.3669]	(0.051)	(0.049)	(0.051)	(0.051)	(0.074)	(0.054)	(0.049)
Percent of immunization	0.653	0.724	0.016	0.017	0.016	0.012	0.027	0.010	0.011
	[0.3664]	[0.3183]	(0.011)	(0.011)	(0.011)	(0.012)	(0.020)	(0.013)	(0.011)
Number of weight checks	2.126	2.227	0.028	0.034	0.028	0.007	0.089	0.019	0.035
	[1.1895]	[1.1464]	(0.040)	(0.040)	(0.041)	(0.043)	(0.061)	(0.041)	(0.040)
Number Vitamin A supplements	1.529	1.507	0.037	0.050	0.042	0.029	-0.043	-0.005	0.045
	[1.1370]	[0.9819]	(0.038)	(0.039)	(0.039)	(0.040)	(0.072)	(0.048)	(0.037)
Percent malnourished	0.168	0.214	-0.006	-0.006	-0.005	-0.005	-0.017	-0.002	-0.006
	[0.3739]	[0.4100]	(0.013)	(0.013)	(0.013)	(0.012)	(0.020)	(0.012)	(0.012)
Age 7–12 gross enrollment	0.948	0.984	-0.005	-0.004	-0.005	-0.004	-0.015	-0.004	-0.005
	[0.2221]	[0.1274]	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.004)	(0.004)
Age 13–15 gross enrollment	0.822	0.891	0.011	0.011	0.011	0.014	0.019	0.005	0.020
	[0.3827]	[0.3125]	(0.014)	(0.013)	(0.015)	(0.015)	(0.025)	(0.015)	(0.015)
Age 7–12 gross attendance	0.904	0.958	-0.001	0.000	-0.001	0.000	-0.001	0.001	-0.002
	[0.2773]	[0.1517]	(0.005)	(0.005)	(0.005)	(0.005)	(0.022)	(0.005)	(0.005)
Age 13–15 gross attendance	0.768	0.872	0.015	0.020	0.011	0.018	-0.009	0.007	0.019
	[0.4125]	[0.3189]	(0.015)	(0.014)	(0.015)	(0.016)	(0.032)	(0.016)	(0.015)
Average standardized effect			0.025	0.0283*	0.025	0.023	0.027	0.031	0.025
			(0.016)	(0.016)	(0.016)	(0.018)	(0.024)	(0.042)	(0.016)
Average standardized effect health			0.0327*	0.0336*	0.0327*	0.025	0.0462*	0.047	0.028
			(0.018)	(0.018)	(0.018)	(0.021)	(0.024)	(0.043)	(0.018)
Average standardized effect educ.			0.009	0.018	0.009	0.019	-0.012	-0.001	0.019
			(0.027)	(0.025)	(0.027)	(0.028)	(0.052)	(0.078)	(0.027)

Appendix Table 2: Regional Heterogeneity, Separately by Wave
Panel A

	Wave II								
Indicator	Java Incentive Treatment Effect	Additional NTT Incentive Treatment Effect	Additional Sulawesi Incentive Treatment Effect	Java Non-Incentive Treatment Effect	Additional NTT Non-Incentive Treatment Effect	Additional Sulawesi Non-Incentive Treatment Effect	Java Incentive Additional Effect	Additional NTT Incentive Additional Effect	Additional Sulawesi Incentive Additional Effect
Main 12 indicators	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number prenatal visits	0.4776*	-0.9764*	0.594	0.059	-1.2910***	-0.331	0.4182*	0.315	0.925
	(0.270)	(0.506)	(0.822)	(0.239)	(0.488)	(0.629)	(0.245)	(0.445)	(0.928)
Delivery by trained midwife	0.022	-0.030	0.1319**	0.0773***	-0.1734**	0.031	-0.0555**	0.1432**	0.1013*
	(0.028)	(0.073)	(0.062)	(0.028)	(0.071)	(0.058)	(0.028)	(0.064)	(0.053)
Number of postnatal visits	-0.4151**	0.479	0.8517**	0.006	-0.4424*	0.281	-0.4207**	0.9216***	0.571
	(0.175)	(0.309)	(0.361)	(0.160)	(0.244)	(0.353)	(0.171)	(0.292)	(0.433)
Iron tablet sachets	0.109	0.062	0.033	0.159	-0.232	-0.3834**	-0.051	0.2936*	0.4166*
	(0.115)	(0.172)	(0.226)	(0.110)	(0.176)	(0.187)	(0.107)	(0.172)	(0.219)
Percent of immunization	0.000	0.045	0.1194*	-0.013	0.062	0.049	0.013	-0.017	0.070
	(0.018)	(0.043)	(0.070)	(0.022)	(0.038)	(0.059)	(0.023)	(0.041)	(0.060)
Number of weight checks	0.1212**	0.028	0.265	0.030	0.056	0.152	0.092	-0.028	0.114
	(0.061)	(0.101)	(0.209)	(0.067)	(0.111)	(0.123)	(0.065)	(0.120)	(0.186)
Number Vitamin A supplements	0.072	-0.3298***	-0.046	0.016	-0.134	0.099	0.056	-0.1957*	-0.145
	(0.060)	(0.123)	(0.189)	(0.070)	(0.124)	(0.168)	(0.072)	(0.114)	(0.243)
Percent malnourished	0.006	-0.0638*	-0.037	0.008	0.001	0.014	-0.002	-0.0643*	-0.051
	(0.018)	(0.039)	(0.055)	(0.017)	(0.038)	(0.043)	(0.019)	(0.036)	(0.060)
Age 7–12 gross enrollment	-0.004	0.019	-0.011	0.002	0.010	-0.009	-0.006	0.010	-0.003
	(0.005)	(0.013)	(0.019)	(0.005)	(0.014)	(0.018)	(0.006)	(0.013)	(0.023)
Age 13–15 gross enrollment	-0.036	-0.030	0.077	-0.018	-0.034	-0.1481**	-0.018	0.004	0.2247***
	(0.026)	(0.047)	(0.049)	(0.030)	(0.046)	(0.059)	(0.031)	(0.046)	(0.063)
Age 7–12 gross attendance	-0.003	0.0243**	-0.024	0.003	0.004	-0.014	-0.006	0.020	-0.010
	(0.006)	(0.012)	(0.021)	(0.006)	(0.013)	(0.015)	(0.006)	(0.013)	(0.023)
Age 13–15 gross attendance	-0.042	-0.030	0.076	-0.028	-0.042	-0.1640***	-0.014	0.012	0.2404***
	(0.027)	(0.047)	(0.051)	(0.030)	(0.047)	(0.063)	(0.033)	(0.048)	(0.066)
Average standardized effect	-0.007	-0.001	0.1434**	0.014	-0.0934**	-0.086	-0.020	0.0924*	0.2296***
	(0.026)	(0.053)	(0.066)	(0.027)	(0.046)	(0.063)	(0.029)	(0.049)	(0.075)
Average standardized effect health	0.029	-0.011	0.1778**	0.036	-0.1194**	0.022	-0.007	0.1082**	0.1561*
	(0.027)	(0.053)	(0.071)	(0.029)	(0.053)	(0.062)	(0.028)	(0.049)	(0.081)
Average standardized effect educ.	-0.078	0.019	0.075	-0.031	-0.042	-0.3017**	-0.047	0.061	0.3764***
	(0.049)	(0.094)	(0.094)	(0.055)	(0.088)	(0.129)	(0.059)	(0.090)	(0.129)

Note: See Notes to Table 5.

Panel B

Indicator	Java Incentive Treatment Effect	Additional NTT Incentive Treatment Effect	Additional Sulawesi Incentive Treatment Effect	Java Non-Incentive Treatment Effect	Additional NTT Non-Incentive Treatment Effect	Additional Sulawesi Non-Incentive Treatment Effect	Java Incentive Additional Effect	Additional NTT Incentive Additional Effect	Additional Sulawesi Incentive Additional Effect
	Wave III								
Main 12 indicators									
Number prenatal visits	0.130	0.136	-0.041	-0.012	-0.055	0.019	0.142	0.192	-0.060
	(0.225)	(0.528)	(0.542)	(0.216)	(0.536)	(0.538)	(0.210)	(0.409)	(0.612)
Delivery by trained midwife	0.019	-0.034	0.001	0.024	-0.107	-0.053	-0.005	0.073	0.054
	(0.024)	(0.062)	(0.051)	(0.024)	(0.071)	(0.058)	(0.023)	(0.057)	(0.071)
Number of postnatal visits	-0.190	0.4893*	0.312	-0.085	0.094	0.223	-0.105	0.395	0.089
	(0.173)	(0.291)	(0.316)	(0.169)	(0.287)	(0.306)	(0.170)	(0.275)	(0.371)
Iron tablet sachets	0.079	-0.035	0.069	0.092	-0.094	-0.189	-0.014	0.059	0.2572*
	(0.072)	(0.149)	(0.147)	(0.086)	(0.164)	(0.162)	(0.085)	(0.142)	(0.153)
Percent of immunization	-0.011	0.076	0.047	-0.008	0.002	0.014	-0.003	0.0741**	0.033
	(0.017)	(0.046)	(0.043)	(0.015)	(0.045)	(0.049)	(0.015)	(0.036)	(0.049)
Number of weight checks	0.1124*	0.143	0.256	0.1831***	0.064	-0.017	-0.071	0.080	0.273
	(0.067)	(0.138)	(0.156)	(0.063)	(0.134)	(0.166)	(0.062)	(0.112)	(0.189)
Number Vitamin A supplements	0.063	0.030	0.111	0.048	-0.176	-0.039	0.015	0.2061*	0.150
	(0.060)	(0.124)	(0.128)	(0.071)	(0.131)	(0.137)	(0.068)	(0.115)	(0.148)
Percent malnourished	0.004	-0.0820**	-0.036	-0.010	-0.0859**	0.025	0.014	0.004	-0.061
	(0.017)	(0.038)	(0.040)	(0.018)	(0.036)	(0.050)	(0.019)	(0.045)	(0.048)
Age 7–12 gross enrollment	-0.005	0.0362***	0.008	-0.003	0.0454***	0.010	-0.002	-0.009	-0.003
	(0.004)	(0.011)	(0.016)	(0.005)	(0.009)	(0.014)	(0.005)	(0.012)	(0.017)
Age 13–15 gross enrollment	0.020	0.009	-0.008	0.025	-0.013	-0.065	-0.005	0.023	0.057
	(0.021)	(0.041)	(0.054)	(0.020)	(0.043)	(0.049)	(0.018)	(0.036)	(0.043)
Age 7–12 gross attendance	-0.007	0.0409**	-0.009	-0.0130**	0.0524***	0.024	0.006	-0.012	-0.033
	(0.007)	(0.016)	(0.024)	(0.006)	(0.014)	(0.017)	(0.007)	(0.016)	(0.025)
Age 13–15 gross attendance	0.027	0.014	-0.022	0.033	-0.029	-0.073	-0.006	0.044	0.051
	(0.022)	(0.044)	(0.055)	(0.021)	(0.045)	(0.047)	(0.020)	(0.037)	(0.043)
Average standardized effect	0.020	0.1060**	0.055	0.0309*	0.022	-0.038	-0.011	0.0845*	0.092
	(0.022)	(0.053)	(0.062)	(0.019)	(0.047)	(0.063)	(0.021)	(0.044)	(0.080)
Average standardized effect health	0.021	0.091	0.093	0.036	-0.027	-0.030	-0.015	0.1173**	0.123
	(0.026)	(0.062)	(0.064)	(0.025)	(0.060)	(0.073)	(0.025)	(0.048)	(0.090)
Average standardized effect educ.	0.018	0.1370*	-0.022	0.021	0.1181*	-0.054	-0.003	0.019	0.032
	(0.035)	(0.073)	(0.086)	(0.033)	(0.067)	(0.078)	(0.033)	(0.067)	(0.083)

Appendix Table 3: Changes over time

Indicator	Wave II			Wave III			Wave III – Wave II		
	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect	Non-Incentive Treatment Effect	Incentive Additional Effect	Incentive Treatment Effect Wave III-Wave II	Non-Incentive Treatment Effect Wave III-Wave II	Incentive Additional Effect Wave III-Wave II
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number prenatal visits	0.300	-0.3559*	0.6556***	0.165	-0.115	0.280	-0.135	0.241	-0.376
	(0.239)	(0.212)	(0.217)	(0.209)	(0.215)	(0.211)	(0.268)	(0.265)	(0.267)
Delivery by trained midwife	0.022	0.025	-0.003	0.023	0.001	0.022	0.001	-0.024	0.025
	(0.028)	(0.027)	(0.025)	(0.025)	(0.027)	(0.027)	(0.030)	(0.030)	(0.031)
Number of postnatal visits	-0.133	-0.035	-0.098	0.034	0.006	0.028	0.166	0.041	0.126
	(0.143)	(0.128)	(0.139)	(0.141)	(0.133)	(0.157)	(0.173)	(0.173)	(0.175)
Iron tablet sachets	0.140	0.064	0.076	0.1356**	0.094	0.042	-0.005	0.030	-0.035
	(0.086)	(0.083)	(0.081)	(0.062)	(0.073)	(0.073)	(0.088)	(0.088)	(0.089)
Percent of immunization	0.028	0.012	0.016	0.017	-0.009	0.025	-0.011	-0.021	0.010
	(0.019)	(0.018)	(0.018)	(0.016)	(0.017)	(0.016)	(0.021)	(0.021)	(0.021)
Number of weight checks	0.1533***	0.058	0.0952*	0.1717***	0.2270***	-0.055	0.018	0.1689***	-0.1505**
	(0.053)	(0.050)	(0.054)	(0.060)	(0.060)	(0.056)	(0.064)	(0.063)	(0.063)
Number Vitamin A	-0.026	-0.012	-0.014	0.088	0.028	0.060	0.114	0.040	0.074
supplements	(0.055)	(0.055)	(0.058)	(0.054)	(0.065)	(0.064)	(0.076)	(0.079)	(0.081)
Percent malnourished	-0.010	0.017	-0.0263*	-0.0374**	-0.0358**	-0.002	-0.028	-0.0525**	0.025
	(0.017)	(0.015)	(0.016)	(0.016)	(0.017)	(0.019)	(0.024)	(0.024)	(0.025)
Age 7–12 participation rate	-0.001	0.003	-0.004	0.0088*	0.0192***	-0.0104**	0.010	0.0162**	-0.006
	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)	(0.005)	(0.008)	(0.008)	(0.007)
Age 13–15 participation rate	-0.0376*	-0.0535**	0.016	0.024	0.017	0.007	0.0616**	0.0707**	-0.009
	(0.019)	(0.022)	(0.024)	(0.018)	(0.019)	(0.016)	(0.026)	(0.028)	(0.028)
Age 7–12 gross attendance	-0.001	0.000	-0.001	0.005	0.008	-0.003	0.006	0.008	-0.002
	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)	(0.008)	(0.009)	(0.010)	(0.009)
Age 13–15 gross attendance	-0.0443**	-0.0680***	0.024	0.0359*	0.022	0.014	0.0803***	0.0902***	-0.010
	(0.020)	(0.023)	(0.024)	(0.019)	(0.020)	(0.017)	(0.027)	(0.029)	(0.029)
Average standardized effect	0.007	-0.030	0.036	0.0716***	0.0515**	0.020	0.0647**	0.0809***	-0.016
	(0.023)	(0.021)	(0.023)	(0.021)	(0.021)	(0.021)	(0.027)	(0.023)	(0.027)
Average standardized effect	0.0468*	0.004	0.0426*	0.0734***	0.040	0.033	0.029	0.038	-0.009
Health	(0.025)	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)	(0.030)	(0.028)	(0.029)
Average standardized effect	-0.0736**	-0.0976**	0.024	0.0679**	0.0737**	-0.006	0.1385*	0.1678***	-0.031
educ.	(0.037)	(0.043)	(0.044)	(0.031)	(0.034)	(0.030)	(0.075)	(0.035)	(0.045)

Notes: See Notes to Table 4. This table restricts analysis to those kecamatans that, according to the randomization, had the same status in both year 1 and year 2. Columns (8) – (10) report the difference between impacts in Wave III and Wave II.