

2011). The observed adverse effect on anaemia in the district-level analysis, on the other hand, may be explained by a higher prevalence of anaemia in target districts before the implementation of the programme, and the low penetration of the programme. Alternatively, the apparent increase in anaemia could reflect an increase in the awareness and diagnosis of anaemia because of the required medical contacts for the children in the programme.

In the analysis of acute malnutrition in children, there was no effect in the individual-level analysis but there were differences in the district-level analysis. This could be explained by programme personnel prioritizing families with children with acute malnutrition for participation. This could result in a baseline difference in nutritional status between Juntos and non-Juntos children, that is only reduced after participation in the intervention. The district level analysis corrects this selection bias. We confirmed that the prevalence of acute malnutrition before implementation was higher in Juntos districts (Table 3). The observed improvement in nutritional status correlates with reported improvements in the quality of food ingested, as found by others (Perova & Vakis 2009b; Segovia 2011) and with micronutrient supplements distributed during health checkups of children at health centres (Dirección General de salud de las Personas 2011). However, there are studies that show that micronutrient supplementation delivered as part of CCT programmes does not have an effect on child nutrition (Attanasio *et al.* 2014).

In our analysis, we did not find any effect on post-partum complications in either analysis. We were expecting that mothers and children would have fewer post-partum complications if they had more deliveries at health centres (Table 2).

One limitation of this evaluation is that we did not have enough baseline data on the prevalence of our outcomes in the district level analysis. There is baseline data for only some of the districts included in Juntos. DHS included a random sample of districts, and therefore a district that is included in one round is not necessarily included in the next round. This resulted in an important reduction in sample size and power for comparisons. Adjusting for these indicators would result in losses of approximately 82% to 85% of the data and increases in the mean bias after propensity score

matching, from 0.5 to 3.42 in the women's database and from 0.8 to 6.6 in the children's database. Although the loss of these respondents reduced our power, it also reflects one of the advantages of the propensity score approach—avoiding extrapolation by limiting analyses to regions of 'common support' and not comparing treated and control observations with very different covariates values. Adjustment by baseline conditions is important, as pre-existing baseline differences could bias our results.

Another limitation of this evaluation is the impossibility of completely removing pre-existing differences between districts. Unlike randomization, propensity score matching only controls for measured differences. Furthermore because of the purposive allocation of the intervention, an observed reduction in a pre-existing difference in an outcome variable could at least be partially explained by regression to the mean if the most severe districts were targeted for the intervention. There was a high degree of selection into this analysis, but that does not necessarily result in selection bias. Nonetheless, if the effect is heterogeneous across different contexts and we have analysed only a subset of observations, then our estimates might indeed lack generalizability or a population-level interpretation. Regression to the mean tends to be an issue when there is measurement error in the indicators. The poorest districts, which were recruited first into the Juntos programme, were stably poor in a way that was more systematic than just a question of measurement error.

A 'difference-in-differences' analysis using another source of data is the natural next step for our study. Data from outpatient clinics routinely collected by the Ministry of Health of Peru by the Health Information System could be a good source for this purpose (Curioso *et al.* 2013).

In concordance with other evaluations of Juntos (Trivelli & Díaz 2010; del Pozo & Guzmán 2011; Escobal & Benites 2012; Perova & Vakis 2012) and evaluations of other CCTs (Carvalho *et al.* 2014; Shei *et al.* 2014), we found good compliance of participants with the programme's participation requirements. We also confirmed the finding of others (Baird *et al.* 2011; Owusu-Addo & Cross 2014; Andersen *et al.* 2015) about the effect of the programme on maternal and child health.