

## El Salvador PBS Activity Evaluation Report

1. Given the evaluation design and evidence in the report, what can we conclude about the impacts of the project?

The evaluation of this dimension of the project is professional and thorough. There is some evidence of a positive impact of the project on the income of dairy producers. It is also clear that the impact varies importantly across sectors, and it seems likely that the timing of the impacts is different depending upon the activity.

2. Can we conclude whether or not this project was successful in achieving its objectives? What would we need to do differently in order to draw conclusions?

The power of the evaluation is relatively low, and so the estimates are not precise. More importantly, we can only see short run impacts in this evaluation. As a consequence, it is difficult to draw very strong conclusions. However, I agree with the thrust of the report: the intervention seems to have had a positive impact on the incomes of dairy producers, and perhaps this is being transmitted through to household income. It is difficult to see much else changing, even with respect to business operations in the other sectors (where the short time frame is less of a concern). Two changes would make stronger conclusions possible: larger sample sizes and a longer time frame for the evaluation. It seems apparent that the evaluation should be designed and interpreted on a sector-by-sector basis. It will be necessary to do a proper cost-benefit analysis separately for each sector for final conclusions to be drawn from this evaluation.

3. What lessons learned/conclusions from the design and implementation of the project and evaluation can be used for future MCC and MCA decisions?

The most important lesson seems to be the dramatic heterogeneity across sectors. The interventions seem to be having remarkably different effects across the different sectors, which is not surprising in itself. But it will be important to understand why the impact seems to be so different across different types of activity.

### Additional Comments

1. As in the other evaluation I reviewed (Armenia) and in the others with which I have some familiarity, there is a very short time frame for impacts to emerge (handicrafts in this evaluation is a partial exception to this). This is a limitation inherent in virtually all randomized roll-out designs and reflects an appropriate goal of not restricting long-term access to programs that are expected to have a strong positive impact.
2. And as in most other clustered evaluation designs, statistical power is quite low and the minimum detectable effect size correspondingly high.

3. The estimated impact on dairy income is remarkably high, particularly given the lack of any evidence of technological innovation among treated dairy producers. This deserves more investigation; what is driving these rapid and large improvements?
4. The lack of generalizability is indeed a limitation. But I think that the framing in the report is misleading. This is *not* a consequence of the small sample size. This is instead because the sample is drawn from (and the randomized design applied only to) the three sectors in the report. Moreover, we have little reason to think that the treatment effects would be similar in any of the other sectors, so we cannot generalize from this report to any of the other activities of the program in El Salvador.
5. The report implies that the standard errors are clustered at the level of the treatment group, but I want to confirm that this is true. And a tiny plea: I find it easier to think about what is going on when standard errors are reported directly, rather than p values, because I can more easily construct confidence intervals in my head.
6. p. 43: is this all the data that was available from FOMILENIO to check on sample construction? In paragraph 3, it would be useful to compare other characteristics (e.g., demographics) to see if the income gap is a consequence of sample selection or of different ways of asking these questions.
7. In section 4A, this is close to data mining: unless we have a priori reason to think that these particular dimensions are the key ones to look at, we should be doing a joint test.
8. Finally, poor balance is the major technical concern on the dairy evaluation. You are taking the correct statistical steps to try to deal with this, but it remains worrisome. It might have been better to ensure balance upfront by matching (Bruhn and McKenzie).