Discussion of:
"Family Planning and Women's Labor Supply: Experimental Evidence from Urban Malawi"
by Mahesh Karra, Daniel Maggio, and David Canning

Discussant:
Tsenguunjav Byambasuren
Cornell University

20th MWIEDC
University of Houston

April 1, 2023

Major comments:

1. Without having separate treatment arms for each component of a treatment package, it is challenging to disentangle the effects of each sub-treatment. But knowing which sub-treatment affects the women's labor market outcomes as well as fertility is important from the policy perspective.

In Section 4.2, you said: "The first, pregnant women, have little need for family planning services while receiving counseling." Does it mean pregnant women are taking up more counseling than others, or they're taking up more family planning services (*per my understanding from my reading of this sentence, free transportation to the Kauma Clinic + financial reimbursement for receiving family planning care at the Kauma Clinic*)? The sub-treatments that are more relevant to pregnant women might be driving the impact of the package because the effects are concentrated among pregnant women, as shown in Figure 1 (panel (b)). The sub-treatments that are more likely to be taken up by postpartum women play less of a role in the overall effects of the package because we don't see any significant effect among postpartum women.

In the next sentence that follows the above one, you said: "However, the counseling is likely more salient and the transport more useful for the second group, postpartum women, who may be actively attempting to limit fertility." So, my understanding from this is postpartum women are more likely to take up 'counseling' + 'free transport to the Kauma Clinic'. If so, these treatments are less effective according to the above logic and the findings in panel (a) of Figure 1, i.e., no significant effects among postpartum women.

Based on these reasonings and related findings, you might be able to say that (i) treatments of counseling and free transport were not strongly effective, and (ii) treatment of financial reimbursement was the most effective treatment among other sub-treatments within the comprehensive package.

As discussed in Section 4.2, in the previous work, you found that most of the effects of MFPS were concentrated among postpartum women. Postpartum women are more readily available to participate in the labor market than pregnant women, so they should drive the effects on labor market outcomes. But this seems not the case because the effects are pronounced among pregnant women. But why? This might imply that the observed effect is due to the sub-treatments taken up

by pregnant women. Particularly, if you compare pregnant and postpartum women with each other, you might isolate the effects of financial reimbursement if the above assumption of treatment take-ups by pregnant and postpartum women is valid. I.e.,

$$\begin{aligned}\text{Pregnant} - \text{Postpartum} &= (\text{free transport} + \text{financial treatment}) - (\text{counseling} + \text{free transport}) \\ &= \text{financial reimbursement} - \text{counseling} \\ &\approx \text{financial reimbursement},\end{aligned}$$

where counseling was found not significant unless it becomes insignificant when it is combined with free transportation treatment (counseling ≈ 0). So, from these, you might be able to say that the effect of comprehensive package is driven by the effects of financial reimbursement.

Also, can you show the first-stage effects (i.e., effects on short-term fertility) for pregnant and postpartum samples separately to make sure that the first- and second-stage heterogeneity results are consistent?

You can do this kind of comparison to identify the effect of each component for both labor market outcomes and fertility (i.e., the first-stage effect). In addition to pregnant/postpartum, you can also think about other sub-groups that might differ in the take-up rate of sub-treatments. For example, if you know the months of pregnancy, you can examine pregnant women who are close to birth and requires frequent visits to hospitals or clinic for examination, which requires transportation. So, such participants might take up free transportation treatment more under the assumption that the Kauma Clinic provides some other childbearing and birth-related services (examinations, some injections, etc.). From the following video, it looks like the Kauma Clinic offers such services not just family planning services: https://www.youtube.com/watch?v=ky2AlqkgD3A

2. Further ideas for heterogeneity

Do you know if treated women discuss what's learned from the family planning services with their husbands? Because fertility and women's labor supply decisions are likely to be mutual decisions based on agreement between husband and wife.

Relatedly, do you also see any increased conflict between wife and husband in a household whose wife was treated? If women refuse to have intercourse to prevent unplanned or close pregnancies or require using protection, husbands could get angry and oppose. All of these could undermine the efficiency of the treatment. So, it might be interesting to estimate the effects heterogeneous by the difference between husband's and wife's age and the difference between husband's and wife's education. The treatment effects might be strong for couples of similar ages or education, who would better understand each other. Perhaps heterogeneity by husband's education individually. The idea here is that the treatment works better if there is an agreement between partners.

3. Weakly significant results and its relation to human behavior literature

The effects are statistically significant but only at the 5% level. What are the potential reasons for that? One of the interpretations could be because it's hard to change human behavior, especially in a 3-year period. So, I think that you can relate your results to the literature on human behavior and behavioral responses to some changes in general. Also, it might be easier to change younger people's

behavior or perspective than older ones. So, I think that you might identify more nuanced results if you do heterogeneity by woman's age, i.e., stronger effects among younger individuals.

Additionally, pregnant women might be taking the treatments more seriously because they received the treatment while going through the problem whereas postpartum women might absorb the treatments or use the treatments less than pregnant women. This might be a behavioral angle of your heterogeneity results.

4. Mechanisms

In addition to the mechanisms that you mentioned in Section 2, I am thinking about if there could be a mechanism that is based on quantity-quality tradeoff. Particularly, I wonder if there is any content related to the quality of the child in the counseling treatment or family planning services in general. The idea is that if women learn that quality is more important than quantity somehow in the counseling sessions or at the family planning care services, this channel might be explaining the negative effects on fertility and positive effects on employment. As the participant understands quality is equally important as quantity, she decided not to have many kids and decided to start working to support the quality of her current children. So, I think that the well-known quantity-quality tradeoff theory by Becker (1960) might be relevant here. Especially, because you mentioned that Maggio et al. (2023) find that the quality of children born directly before the intervention, this literature seems relevant. Or do you think that children's quality is improving as the parents start working and earning income and invest in their children?

Since the contents of the free counseling and family planning care or services at the clinic are not provided, it is hard to know what treatments the participants took. So, if it is OK, the readers would appreciate seeing some detailed descriptions of the contents of the counseling and family planning services in the Appendix.

5. Time-use results

The results on women's time use are quite puzzling. It is the women's time use on workdays, not weekends, right? What is the exact question being asked? Although you said that it's only for a short window of 24 hours, I assume that the respondent should have answered the question thinking that they've been asked what portion of their time is used in a regular workday on average. So why their time spent on their job does not increase when they start working? From the results on employment, it looks like those women who started working seem to transition from unemployment because other employment types do not decrease. So, there should be an increase in the time used for work by a woman who switched from unemployed to employed. But there is not.

Perhaps heterogeneity analysis for time-use regressions might unravel some stories behind the insignificant impacts on time use. You can do the same three heterogeneities as used in the women's employment regressions. Particularly, I consider that employed vs. not employed at baseline might be a more important heterogeneity for the time-use regressions because time use of not employed at baseline women is expected to rise. Also, since the estimates for pregnant women are highly significant and stronger than full-sample women results, you might find some significant change in time use among pregnant women.

6. Men's results

Results on men's employment show that husbands are not sitting back when their wives start working (i.e., no deleterious effects). But there was no discussion of why men's employment increased. So, I think that it can be interpreted based on intra-household bargaining power. Men start working when their wives start working to not get behind and maintain their power in the household in terms of generating income.

Also, it might not make sense. But I wonder if those men who start working were husbands of women start getting employed. Can you track those men who start working and match them with their wives and show if both partners started working during the study period?

7. Robustness checks

   a) Robustness check of baseline results to logit/probit regressions given that both the outcome and treatment variables are binary.
   b) Robustness check of heterogeneity results to using interaction terms instead of sample splitting which suffers statistical power.

Minor comments:

1. As stated in the Introduction section, the paper contributes to two strands of literature. For both contributions, you said that there are few other studies on the exact topic using similar methods. Particularly,
   a) "However, there is a relative dearth of experimental evidence on the topic."
   b) "However, due to challenges associated with measuring downstream causal effects, there is a relatively small amount of evidence linking family planning programs to medium and long-term outcomes."

   Does this mean there is no other study doing the same thing and this paper is the first? Or a few other experimental papers are doing a similar thing? If so, what are those other few papers and what are the main distinctions of this paper from those? The thing that I wonder was that is this the first study to use RCT to look at the impact of family planning programs on labor market outcomes.

   However, I find that there is another paper, Barham et al. (2021) who examine the effects of family planning programs on labor market outcomes in Bangladesh. So, I think it's better to discuss this paper in the Introduction section and clearly describe the distinctions and innovations of the present paper leading to different findings. Is it only because of context difference?

2. When I first read the abstract, I thought about a collective labor supply model by Cherchye, Rock, and Vermeulen (2012) https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.102.7.3377. I think your paper is not strongly related to this literature, but I think that there might be some link as you're talking about fertility, labor supply, and time use.

3. It was not clear what exact labor market outcomes you looked at until I reach the middle of the paper (or Section 4 on page 8). I kept wondering if they were employment, wage, employment

transition, search, etc. until I reached that point. So, it might be helpful for readers what are the labor market outcomes you looked at sooner than that, perhaps somewhere in the Introduction section.

4. Can you provide a little bit of details about the treatment of financial reimbursement for family planning care at the Kauma Clinic? To receive a reimbursement, do women have to show you receipts or prescriptions from the clinic with dates and times? Is it possible to fake them or multiply one copy to receive reimbursements for services that they didn't get? It is also possible that participants ask the clinic people to make a form that says they visited and received services with some arrangement with the clinic people. So, I wanted to have some discussion about this specific sub-treatment, which seems like driving the main results, because the other two sub-treatments can be easily monitored or they seem more direct. Relatedly, do you see any unusual number of visits by participants since one can redeem multiple visits? For example, multiple visits in short periods.

5. Is running an enterprise the same as being self-employed? So, does an outcome in Column (1) of Table 1 include outcomes in Columns (2) and (3)? Clarifying this might be helpful for readers. What if you define an alternative employment variable, including agricultural work as well? If the coefficient is still positive and significant, then it confirms that these women transition from unemployment to employment.

6. Footnote 2 might not be correct because you have labor force participation, unemployment, and husband's labor market outcomes.

7. In Figure 1, I assume that those CIs are 95% and based on heteroskedasticity-robust SEs. But a short note under the figure providing that information should not be repetitive.