It depends who's asking? A multi-country field experiment on surveyor identity and response bias[*]

Rachel Cassidy[†]   Jonathan Lain[‡]   Simone Lombardini[§]

July 13, 2018

PRE-ANALYSIS PLAN

1   Introduction

Potential response bias in household surveys is a concern for studies in many contexts. Recent literature has considered the role of experimenter demand in lab and lab-in-field experiments (Zizzo, 2010; de Quidt et al., 2017), including the impact of the enumerator's race or ethnicity and gender (Cilliers et al., 2015; White et al., 2016). Meanwhile, Stecklov et al. (2017) show that whether enumerators pay respondents matters for participation rates and responses in household surveys, with respondents who receive a flat fee appearing to underreport their expenditures on luxury items and their ownership of unverifiable assets. However, there is little evidence on whether respondents' perceptions of the organisation conducting the survey – for example, whether it is an NGO conducting projects in the region, or rather an independent research institute – influences response rates and the accuracy of data collected. Anecdotally, it appears that many researchers have often made the choice of which partner organisation(s) to highlight at the start of a survey on a project-by-project basis.[1] Yet the identity of the organisation responsible may have differential effects on respondents' trust, social desirability concerns, and beliefs that their responses may influence future program delivery. Moreover, this

---

[†]Corresponding author. Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, United Kingdom. Email: rachel.cassidy@ifs.org.uk

[‡]Organisation for Economic Co-Operation and Development.

[§]Oxfam Great Britain.

[1]In the case of Oxfam's Effectiveness Reviews, this choice is made in adherence to Oxfam's Responsible Data Policy and its standards for informed consent, which were reviewed and updated following the introduction of the new EU GDPR legislation in June 2018.

effect may differ depending on whether communities or respondents have been the recipients of an intervention, biasing estimates of treatment impacts. Evidence on the potential biases created by making salient one particular type of organisation is therefore important, both in weighing the evidence from past studies, and in informing ongoing debates about best practices for collecting informed consent.

Mindful of these concerns, and as part of its ongoing attempts to improve data quality, Oxfam Great Britain (henceforth "Oxfam") agreed to implement a large-scale experiment in 2016. The purpose was to test whether individuals' responses are systematically influenced by the type of organization that they believe to be conducting the study. The research team embedded an experiment in Oxfam's evaluation of eight different interventions in eight different countries. We randomised at the respondent level whether we made salient that the survey was on behalf of an international NGO, a local NGO, or a local independent survey company.

This pre-analysis plan provides full details of the sampling and experimental design. It then specifies how we will estimate whether the experimental treatment arms affect respondents' reporting of key outcome variables. To explore whether this might create bias in estimating treatment effects from interventions, we will also test whether the effects of the experimental treatment arms differ across individuals from Oxfam's project intervention groups compared to individuals from Oxfam's comparison groups. These groups are identified by Oxfam via quasi-experimental methods for its own project Effectiveness Reviews. At the time of writing, the data from all eight surveys have been used by Oxfam for assessing the impact of their interventions via their Effectiveness Reviews.[2] This analysis involved comparing Oxfam's project intervention groups with their quasi-experimental comparison groups, keeping data from all of our experimental treatment arms together. No individual has examined the data separated by our experimental treatment arms.

2   Context and sampling

The experimental sample consists of all 5,641 individuals interviewed during the course of eight Effectiveness Reviews (quasi-experimental impact evaluations) carried out by Oxfam in 2016. These evaluations took place across eight countries: Honduras, Indonesia, Kenya, Niger, South Sudan, Tanzania, Tunisia, and Zambia. The interventions evaluated fall under Oxfam's thematic areas of Resilience (Kenya, Zambia), Livelihoods (Niger, South Sudan, Tanzania) and

---

[2]https://policy-practice.oxfam.org.uk/our-approach/monitoring-evaluation/effectiveness-reviews.

Women's Empowerment (Honduras, Indonesia, Tunisia). In each evaluation, Oxfam surveyed a random sample of individuals in its project intervention group and a sample of individuals identified as a comparison group. The intervention group consisted of either: i) a random sample from the list of all intervention participants, in countries where the intervention under evaluation was rolled out at the individual level; or ii) a random sample of all households from intervention communities (with one respondent being interviewed per household), in countries where the intervention was rolled out at the community level. Methods for selecting a comparison group varied depending on the context of each evaluation. Typically, nearby communities were selected with similar characteristics to those where the project was implemented, and individuals from those communities were selected on the basis of the intervention's targeting criteria. In the remaining cases, comparison communities or respondents were those who had been selected for participation in a later phase of the intervention.[3] The distribution of respondents across countries, thematic areas and intervention group/comparison group status is shown in Table 1.

Table 1: Sample of respondents by country and thematic area

Thematic area	Country	Intervention respondents	Comparison respondents	Total
Resilience	Kenya	289	563	852
Resilience	Zambia	282	545	827
Livelihoods	Niger	300	404	704
Livelihoods	South Sudan	316	385	701
Livelihoods	Tanzania	229	468	697
Women's Empowerment	Honduras	237	302	539
Women's Empowerment	Indonesia	295	505	800
Women's Empowerment	Tunisia	230	291	521
Total		2,178	3,463	5,641

3 Experimental design

Our experiment was added as an extra component to Oxfam's surveys. Our treatment randomised at the respondent level which organisation(s) were mentioned at the start of the interview, when explaining the research objectives and gaining participants' consent. In each case,

---

[3]In its Effectiveness Reviews, Oxfam also uses propensity-score matching to provide additional confidence that the two groups — intervention and comparison — do not differ based on a number of observable recalled baseline characteristics (Hutchings, 2011, 2014; Oxfam, 2016). Typically the number of comparison respondents is larger than the number of intervention respondents, to help facilitate matching of the two groups.

both the name of the organisation and a short qualifying description were given, for example:

- Oxfam treatment arm:"I am conducting a survey on behalf of Oxfam. Oxfam is an international organisation working to develop long-lasting solutions to poverty and promote campaigns for social change"; or

- Local consultant treatment arm: "I am conducting a survey on behalf of <<consultancy name>>. <<Consultancy name>> is a consultancy company from <<country name>> specialised in collecting socio-economic information."

The treatment arms implemented in each country were as follows:

- In Kenya, Niger, South Sudan, Tanzania, Honduras and Indonesia, the treatment arms were $\{LocalConsultant, Oxfam\}$ each with probability 50%.

- In Zambia, the set of treatment arms was expanded to $\{LocalConsultant, Oxfam, LocalPartner\}$ each with probability one-third.

- In Tunisia, excluding Oxfam's name was not feasible due to the political economy of the setting; thus the treatment arms were $\{Oxfam, Oxfam + LocalPartner\}$, each with probability 50%.

The randomisation was built into the software on the mobile devices that were used to conduct the surveys. The random number generator and assignment to treatment arm occurred at the moment that the enumerator started a new survey, thus prior to the collection of any covariates. Each respondent within a given country faced the same probability of receiving each treatment arm. Thus altogether, this was equivalent to pure randomisation stratified by country. Table 2 shows the number of observations in each country based by treatment arm.

The specific protocol was as follows. When starting the survey, the enumerator was randomly directed to a consent form page that contained only the names of the organisation(s) to be mentioned in that treatment arm. Enumerators were informed in advance that this would occur, and were carefully trained to mention the name of the organisation(s) displayed on the page and that organisation(s) only, unless explicitly asked by the respondent. They were told that this was part of attempts to improve data quality, but not given information about the specific research question or hypotheses. To increase exposure to the treatment, the relevant organisational logo was shown on the screen when respondents were asked to tap to consent

Table 2: Sample of respondents by country and treatment arm

Country	Consultant	Oxfam	Oxfam+Local Partners	Local Partners	Total
Honduras	274	265	0	0	539
Indonesia	415	385	0	0	800
Kenya	424	428	0	0	852
Niger	354	350	0	0	704
South Sudan	308	393	0	0	701
Tanzania	352	345	0	0	697
Tunisia	0	268	253	0	521
Zambia	271	289	0	267	827
Total	2,438	2,723	285	267	5,641

participating in the survey; see Figure 1.[4] The rest of the survey interview then proceeded as normal. At the very end of the interview, respondents were asked two questions to elicit their perceptions of i) which organisation(s) were involved in the survey and ii) the survey purpose.
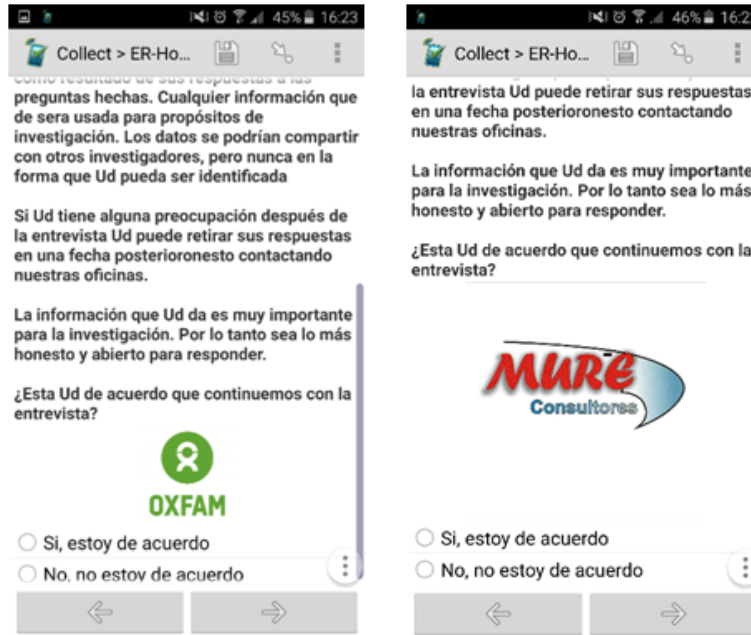
Figure 1: Consent form — Oxfam treatment arm

To ensure standards for ethical research were met — specifically, ensuring informed consent and avoiding any deception of respondents — the following key measures (among others) were taken. First, the consent form explicitly stated that data would be used for research and evaluation purposes. Second, respondents were told that in the case that data would be shared

[4]Two exceptions due to local political sensitivities were as follows: i) in Indonesia, Oxfam's logo was shown in the Oxfam treatment arm, but the local consultant's logo was not shown in the local consultant treatment arm; ii) Tunisia, no logos were shown for any treatment arms.

Figure 2: Logos on consent forms in Honduras — Oxfam treatment arm (left) and local consultant arm (right)

with other organisations, it would not be shared in a way in which any household could be identified.[5] Third, when the field supervisors contacted village authorities to arrange permission to enter each village, they were trained to mention the names of all organisations involved. Fourth, if a respondent asked for more information on the organisation(s) conducting the survey (e.g. asking if a particular organisation was involved, or if there were any other organisations involved beyond the one mentioned during the consent form), enumerators were carefully trained to truthfully mention all organisations involved.[6] The treatment can therefore be interpreted as making one particular organisation salient to each respondent at the start of the interview, and framing the interview as being connected to that organisation in particular, rather than fully restricting information about which organisations were involved in the survey process.

---

[5]Oxfam shared only anonymised data with the research team.

[6]This was outlined explicitly in the documents shared and agreed with the survey firms in each country. In Kenya, for example, the Memorandum of Understanding with the consultants leading the team of enumerators contained the following text: "As part of its constant efforts to improve the quality of its survey work, Oxfam is hoping to use this Effectiveness Review to scientifically test whether respondents' perceptions of who is conducting a particular survey, influence the answers provided... To test the presence and size of such effects, we wish to randomise the consent forms that used to begin the interviews. Some of the consent forms will mention that the survey is being conducted by Oxfam, whilst other consent forms will mention that the survey is being conducted by [name of consultancy firm]. Please note that we are not trying to deceive respondents, since both of these cases are true. **Also, although enumerators should only mention one particular organisation, if they are asked directly whether Oxfam or [name of consultancy firm] are involved, they must answer truthfully that both organisations are involved.**"

4   Data

The data come from the eight different endline surveys, one per country, that Oxfam uses to evaluate impact of the interventions. Some core survey modules are common across all surveys; other modules and additional questions are common only to surveys in the same thematic area (i.e. Resilience, Livelihoods, Women's Empowerment), or are specific to a particular intervention. Appendix A.1 describes in full the construction of the variables specified in this pre-analysis plan from the raw survey variables. To ensure robustness to outliers, for the main analyses we will winsorize at the 95th percentile any continuous variables that were unbounded by the tablet survey template.

4.1   Attrition

Since the experiment takes place within a set of cross-sectional endline surveys without baseline counterparts, there is no panel dimension to the data, and hence no scope for attrition in the classic sense. However, there are two ways in which respondents may appear in the sampling frame but not the actual data:

1. They are in the sampling frame but cannot be contacted for interview

   In this case the respondent drops out of the study before assignment to one of our treatment arms, which occurs when the enumerator opens the survey at the start of the interview; thus this is not a concern for our analysis.

2. They refuse to consent to participate in the survey

   In principle this could be affected by the treatment arm that a respondent is assigned to, e.g. if respondents who are told that the survey is being conducted on behalf of a local consultant are less likely to agree to participate. This has the potential to create selection bias in our analysis of the impact of the treatment arms on responses to survey questions. Consent is also an outcome variable of interest, insofar as participation rates are an important factor to consider when making the choice between which organisation(s) to mention at the start of a survey interview.

   Refusal to participate in the study was registered by enumerators after reading the consent form. We will therefore regress consent to participate in the survey on a dummy for each

of the treatment arms, with the Oxfam treatment arm as the excluded category. We will include country fixed effects, since treatment was effectively stratified by country, and will cluster standard errors at the individual level, since this was the level of randomisation. The estimating equation will be the following logit specification:

$$Pr(y_{ic} = 1 | local\_consultant_{ic}, oxfam\_local\_partner_{ic}, local\_partner_{ic}, \mu_c)$$
$$= \Lambda(\beta_0 + \beta_1 local\_consultant_{ic} + \beta_2 oxfam\_local\_partner_{ic} + \beta_3 local\_partner_{ic} + \mu_c)$$
$$(1)$$

for individual i in country c. y_{ic} is a dummy variable indicating whether the individual consented to be interviewed, i.e. to become a respondent. $local\_consultant_{ic}$ is the *local consultant* treatment arm; $oxfam\_partner_{ic}$ is the *Oxfam and partners* treatment arm (Tunisia); $partner_{ic}$ is the *partners only* treatment arm (Zambia); and the *Oxfam* treatment arm is the omitted category. μ_{ic} is a country fixed effect. We will cluster standard errors at the individual level, given that this was the level of randomisation. We will report the coefficients $\beta_1, \beta_2, \beta_3$, and in each case both the p-value of the test that the coefficient is different from zero. If we observe that any of the treatment conditions is a significant predictor of non-consent, we will estimate Lee bounds on the main treatment effects estimations listed below.

4.2   Balance

All survey variables are measured after exposure to treatment, i.e. after a particular organisation(s) has been made salient to the respondent during the introduction of the survey and signing of the consent form. Given that the treatments are designed to induce conditions under which reporting bias might differ, any variable which is vulnerable to misreporting might appear imbalanced across treatment arms, but this may be evidence of a treatment effect rather than imbalance in the treatment assignment. We therefore limit balance testing to variables which are highly unlikely to be vulnerable to misreporting, because they are close to perfectly verifiable by the enumerator. Taking the full consenting sample, i.e. the full sample for whom we have data other than the treatment assignment and consent/non-consent variable, we will therefore test for balance on the following variables:

Balance variables:

- Gender of the respondent (*gender_int*)

- Housing (*modern_walls_now, modern_roof_now, modern_floor_now*)[7]

- Respondent is from the majority ethnic group (*mainethnicgroup_int*) – in countries where collected

- Individual is in the intervention group vs the comparison group (*intervention*)

We will estimate Equation 1 as above for each of the binary balance variables, and the following Ordinary Least Squares specifcation for each of the continuous variables:

$$y_{ic} = \beta_0 + \beta_1 local\_consultant_{ic} + \beta_2 oxfam\_local\_partner_{ic} + \beta_3 local\_partner_{ic} + \mu_c + \epsilon_{ic} \quad (2)$$

Standard errors will be clustered at the individual level, given that this was the level of randomisation. We will report the following statistics:

- For each characteristic, the magnitude of β_1 and the p-value of the test that $\beta_1 = 0$. This corresponds to a test that assignment between the Oxfam treatment arm and the local consultant treatment was balanced for that balance variable. Note that identifying variation comes from all countries except Tunisia, where this treatment arm was not implemented.

- An F-test that the vector of β_1 coefficients estimated for each of the balance variables are jointly equal to zero. This is the test we will focus on to assess balance for assignment to the Local Consultant treatment arm as compared to the Oxfam treatment arm (Honduras, Indonesia, Kenya, Niger, South Sudan and Tanzania), since it takes into account the fact that we are testing for balance across multiple balance variables which may spuriously appear imbalanced.

- For each characteristic, the magnitude of β_2 and the p-value of the test that $\beta_2 = 0$. This corresponds to a test that assignment between the Oxfam treatment arm and the Oxfam and Partners treatment arm was balanced on that balance variable. Note that identifying variation comes from Tunisia only, as the Oxfam and Partners treatment was only implemented in Tunisia.

---

[7] In Honduras and Tunisia, interviews took place within community centres rather than respondents' homes. We may therefore observe imbalance on these housing variables in those countries, if there was systematic misreporting by treatment arm.

- An F-test that the vector of β_2 coefficients estimated for each of the balance variables are jointly equal to zero. This is the test we will focus on to assess balanced treatment assignment in the Tunisian experiment.

- For each characteristic, the magnitude of β_3, the p-value of the test that $\beta_3 = 0$, and the p-value of the test that $\beta_1 = \beta_3 = 0$. The latter corresponds to a test that assignment between the Oxfam treatment arm, the Local Partners arm and the Partners (only) treatment arm was balanced on that balance variable. Note that identifying variation comes from Zambia only, as the Partners treatment was only implemented in Zambia.

- An F-test that the vector of $\beta_3 = 0$ coefficients estimated for each of the balance variables are jointly equal to zero; and an F-test that the vectors $\beta_1 = \beta_3 = 0$. This latter statistic is what we will focus on to assess balanced treatment assignment in the Zambian experiment.

5 Estimating equations

Our main research question is whether the identity of the organisation named as conducting a survey systematically influence participants' responses. To answer this, we will test for effects of our consent form treatments on the following families of outcome variables (variable names in italics). We label the outcomes as "primary" and "secondary" depending on the strength of our prior that the salience of a particular organisation is likely to bias responses for that variable.[8] For binary outcomes, we will again estimate Equation 1. For continuous outcomes, we will again estimate Equation 2. Within each family, we will report both the p-values and the False Discovery Rate-adjusted q-values taken across all variables within that family of outcomes (Anderson, 2012).

Primary outcomes

1. Subjective wealth

    - Subjective relative wealth (*subwealth_eastafrica_year* for Kenya and Tanzania, *subwealth_now_all* for the remaining countries)[9]

    - Number of days household could survive on savings (*saving_alldays*, Kenya and Zambia only)

---

[8]The terms "primary" and "secondary" hence do not indicate in any way the importance of these outcomes as perceived by Oxfam.

[9]These variables are on a 1-10 scale, but will be treated as continuous.

2. "Objective" wealth

  - Livestock (*livestock_index*, Kenya, Niger, South Sudan, Tanzania and Zambia only)

  - Assets index (*assets_index*, Kenya, Niger, South Sudan, Tanzania and Zambia only)

  - Oxfam's wealth index (*wi_unclean_norm_now*)

3. Reported receipt of programme support

  - Receipt of training (*training*)

  - Receipt of inputs (*inputs*, Niger, South Sudan and Tanzania only)

Secondary outcomes

4. Consumption

  - Household consumption adult-equivalent expenditure per day (*cons_total_pd_paeu*)

5. Demographics and typical programme eligibility criteria

  - Household size (*hhnumber*)

  - Number of dependents (*dependentsnum*)

  - Household head ability to work (*fitwork_hhh*)

6. Women's empowerment

  - Violence (*violence_you*, *violence_wclose*)

  - Oxfam's women's empowerment index (*wei*)

7. Resilience

  - Oxfam's resilience index (*res_bri_basic*)

5.1   Heterogeneity

We plan to report heterogeneity of the effects above along the following dimensions.

Intervention vs comparison groups: For the full sample, we will test whether the treatment effects estimated for the above outcome families are stronger among project intervention or comparison respondents. To do so we will estimate Equations 1 and 2 for binary and continuous outcome variables respectively, adding a dummy for whether the respondent is from the intervention group, and its interaction with each of the treatment dummies. Again, we will report the estimated coefficients and their p-values, but also the FDR-adjusted q-values taken across all variables within each family of outcomes.

Country-by-country: As well as reporting the results for the full sample, we will also report each of the above regressions for the above outcome families estimated separately by country. When estimating separately by country, we will drop the OP treatment dummy for all countries except in all countries except Tunisia, since this treatment arm was only implemented there. We will also drop the LC treatment dummy in the Tunisian regression, since that treatment arm was not implemented there. Similarly, we will drop the P treatment variable for all countries except Zambia, since this treatment arm was only implemented there. Within each country, we will again correct p-values within each family of outcomes and report FDR-adjusted q-values alongside p-values.

6 Robustness

Manipulation check: We will use the exit question "which organisation(s) do you think is (are) responsible for conducting this survey?" to check whether the treatment appeared to move participants' perceptions of which organisation was responsible for the survey. To do so, we will estimate the same reduced-form specification as above in Equation 1, with the following outcome variables as the dependent variables:

(N.B. multiple, unranked responses were possible, and no prompting was given)

- Surveyor is an international NGO (*surveyorg _ingoall*)
- Surveyor is a local NGO (*surveyorg _localngoall*)
- Surveyor is a consultant (*surveyorg_consultant*)

In each case, we will report the estimated coefficients from Equation 1 and the p-values of the tests that they are equal to zero.

We predict that:

a The Local Consultant treatment will have a negative impact on the outcome "surveyor is an international NGO" (all countries excluding Tunisia);

b The Local Consultant treatment will have a negative impact on the outcome "surveyor is a local NGO" (all countries excluding Tunisia);

c The Local Consultant treatment will have a positive impact on the outcome "surveyor is a consultant" (all countries excluding Tunisia);

d The Partner treatment will have a positive impact on the outcome "surveyor is a local NGO" (Zambia only);

e The Oxfam and Partners treatment will have a positive impact on the outcome "surveyor is a local NGO" (Tunisia only).

For each country, if we do not observe at least one of the predicted effects for that country to be significant at the 10% level or less, then we will re-run the main estimations for outcome families 1-5 dropping that country.

Similarly, to provide evidence on the possible mechanisms behind any observed effects, we will examine responses to the exit question "Do you happen to remember what is the primary purpose of this survey?". We will estimate Equation 1 for the following variables:

(N.B. these variables are not mutually exclusive)

- Survey purpose is evaluation ($surveypurpose\_evaluation$)
- Survey purpose is targeting ($surveypurpose\_targeting$)
- Survey purpose is research ($surveypurpose\_research$)

In each case, we will report the estimated coefficients from Equation 1 and the p-values of the tests that they are equal to zero.

Controls: We will report whether each of the estimated treatment effects are robust to the inclusion of covariates. The covariates will be identified by including all of the balance variables specified above into a post-double-LASSO procedure.

References

S. Alkire, R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymour, and A. Vaz. The women's empowerment in agriculture index. *World Development*, 52:71–91, 2013.

M. L. Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 2012.

D. Bishop and K. Bowman. Still learning: a critical reflection on three years of measuring women's empowerment in Oxfam. *Gender & Development*, 22(2):253–269, 2014.

J. Cilliers, O. Dube, and B. Siddiqi. The white-man effect: How foreigner presence affects behavior in experiments. *Journal of Economic Behavior & Organization*, 118:397–414, 2015.

J. de Quidt, J. Haushofer, and C. Roth. Measuring and Bounding Experimenter Demand. *National Bureau of Economic Research Working Paper*, 2017.

A. Deaton and S. Zaidi. *Guidelines for constructing consumption aggregates for welfare analysis*, volume 135. World Bank Publications, 2002.

D. Filmer and L. H. Pritchett. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, 38(1):115–132, 2001.

R. Fuller and J. Lain. Measuring Resilience: Lessons learned from measuring resilience in Oxfam's large-N Effectiveness Reviews. *Oxfam Discussion Papers*, 2015. URL https://policy-practice.oxfam.org.uk/publications/measuring-resilience-lessons-learned-from-measuring-resilience-in-oxfams-large-583601.

M. Grosh and P. Glewwe. Designing household survey questionnaires for developing countries. *World Bank Publications*, 2000.

C. Hutchings. Can We Obtain the Required Rigour without Randomisation? Oxfam GB's Non-Experimental Global Performance Framework. *International Initiative for Impact Evaluation (3ie) Working Papers*, (13), 2011. URL http://www.ipdet.org/files/Publication-Can_we_obtain_the_required_rigour_without_randomization.pdf.

C. Hutchings. Balancing accountability and learning: a review of Oxfam GB's global performance framework. *Journal of Development Effectiveness*, 6(4):425–435, 2014.

S. Lombardini and K. McCollum. Using internal evaluations to measure organisational impact: a meta-analysis of Oxfam's women's empowerment projects. *Journal of Development Effectiveness*, 10(1):145–170, 2018.

Oxfam. How are effectiveness reviews carried out? *Oxfam Online Publications*, 2016. URL https://policy-practice.oxfam.org.uk/publications/how-are-effectiveness-reviews-carried-out-594353.

D. E. Sahn and D. Stifel. Exploring alternative measures of welfare in the absence of expenditure data. *Review of income and wealth*, 49(4):463–489, 2003.

G. Stecklov, A. Weinreb, and C. Carletto. Can incentives improve survey data quality in developing countries?: results from a field experiment in India. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2017.

A. White, A. Strezhnev, C. Lucas, D. Kruszewska, and C. Huff. Investigator Characteristics and Respondent Behavior in Online Surveys. *Working Paper*, 2016.

D. J. Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98, 2010.

A Appendices

A.1 Variables

Variable	Countries	How constructed
country	All	set to country name
hhid	All	unique respondent identifier within country
intervention	All	= 1 if respondent in intervention group, = 0 if in comparison group
oxfam	All	Oxfam treatment arm (assigned by random number generation in tablet)
local_consultant	All	Local Consultant treatment arm (assigned by random number generation in tablet)
local_partner	Zambia	Local Partner treatment arm (assigned by random number generation in tablet)
oxfam_local_partner	Tunisia	Oxfam and Local Partner treatment arm (assigned by random number generation in tablet)
consent_yn	All	= 1 if respondent consented to interview, either outright or after queries or re-explanation; = 0 if not
gender_int	All	= 1 if respondent male, = 0 if female
modern_walls_now	All	= 1 if wall material modern/improved, = 0 otherwise
modern_roof_now	All	if roof material modern/improved, = 0 otherwise
modern_floor_now	All	if floor material modern/improved, = 0 otherwise
mainethnicgroup_int	All except Tunisia	if respondent belongs to main ethnic group in survey sample, = 0 otherwise
surveypurpose_evaluation	All except Tanzania	= 1 if respondent's unprompted, open-ended answer to "Do you remember what is the main purpose of this survey?" was coded by the enumerator as including "To evaluate a past government program", "To evaluate a past non-governmental program", or "To decide whether a program should continue or not" (enumerators could code as many options as they deemed applicable); = 0 otherwise.

surveypurpose_targeting	All except Tanzania	= 1 if respondent's unprompted, open-ended answer to "Do you remember what is the main purpose of this survey?" was coded by the enumerator as including "To decide which households or communities most need a future governmental program" (enumerators could code as many options as they deemed applicable); = 0 otherwise
surveypurpose_research	All except Tanzania	= 1 if respondent's unprompted, open-ended answer to "Do you remember what is the main purpose of this survey?" was coded by the enumerator as including "To understand the lives of people in this community" (enumerators could code as many options as they deemed applicable); = 0 otherwise
subwealth_eastafrica_year	Kenya, Tanzania	"Thinking about all the sources of income that members of your household have (including selling of assets/livestock, remittances, and NGO/government transfers) what was your approximate total household income for the past 12 months?" (local currency units)
subwealth_now_all	Honduras, Indonesia, Niger, South Sudan, Tunisia, Zambia	"On a scale of one to ten, where one is the poorest household in your local community and ten is the wealthiest household in your local community, how wealthy do you consider your household to be?" (number 1-10)
training	All	= 1 if respondent reports the household received training, = 0 otherwise
inputs	Niger, South Sudan, Tanzania	= 1 if respondent reports the household received inputs (seeds, tools), = 0 otherwise
saving_alldays	Kenya, Zambia	"If you had an emergency now and had to stop working, how many days could you support yourself and your household on the money and other goods (e.g. food) you have saved?" (number of days)

livestock_index	All	Sum of: cattle_now camels_now horses_now donkeys_now sheep_now goats_now pigs_now and poultry_now (each variable is the absolute number of animals)
assets_index	All	Sum of: moto_now, tv_now, fridge_now, gasstove_now, mobilephone_now, computer_now, jerrycan_now, carrier_now, hoe_now, machetescythe_now, rake_now, axe_now, waterpump_now, wheelbarrow_now, cart_now, plough_now, harrow_now, tractor_now, fishingnet_now, fishinghook_now, wallclock_now, watch_now, mat_now, stool_now, chair_now, table_now, mattress_now, lamp_now, generator_now, cookingpot_now, iron_now, radioplayer_now, solar_now, bicycle_now, boat_now, file_now, watertrough_now, bed_now, sleepinghide_now, box_now, sewingmachine_now, jewellery_now, tradclothes_now, washingmachine_now, heating_now, airconditioning_now, microwave_now, sofa_now, dvdvideoplayer_now, and fan_now (each variable is the absolute number of assets)
wi_unclean_norm_now	All	Oxfam's wealth index: the housing variables, livestock variables and assets variables listed above are taken together, along with land owned in Honduras and Indonesia only; the first principle component is estimated; this component is then normalised as a z-score by country. As proposed by Filmer and Pritchett (2001) and Sahn and Stifel (2003).
cons_total_pd_paeu	Niger, South Sudan, Tanzania	Oxfam's calculation of consumption per adult equivalent unit per day, with 1 child=0.33 adults, using the methodology of Grosh and Glewwe (2000) and Deaton and Zaidi (2002).
hhnumber	All	number of individuals in the respondent's household

dependentsnum	All	number of dependents ($children < 16 + adults > 65 + 16 \leq adultswithdisability \leq 65$ in the household
fitwork_hhh	All	$= 1$ if respondent reports household head fit to work, $= 0$ otherwise
violence_you	Honduras, Indonesia, Tunisia	"In the past 12 months, do you think someone may have done these things to you?..." $= 1$ if yes to one or more of: "threaten to hurt or harm her or someone she cares about"; "insult her or make her feel bad about herself"; "push her, shake her, slap, punch her or thrown something at her"; "threaten or attack her with a knife, gun or other weapon"; $= 0$ otherwise.
violence_wclose	Honduras, Indonesia, Tunisia	"In the past 12 months, do you think someone may have done these things to a woman close to you?..." as above for *violence_you*
wei	Honduras, Indonesia, Tunisia	Oxfam's women's empowerment index (Bishop and Bowman, 2014; Lombardini and McCollum, 2018)
res_bri_basic	Kenya, Zambia	Oxfam's resilience index (Fuller and Lain, 2015)
surveyorg_ingoall	All except Tanzania	$= 1$ if respondent's unprompted, open-ended answer to "Which organisation or organisations do you think are conducting this survey?" was coded by the enumerator as including any international NGO (including Oxfam), by name or described as a category (enumerators could code as many options as they deemed applicable); $= 0$ otherwise
surveyorg_localngoall	All except Tanzania	$= 1$ if respondent's unprompted, open-ended answer to "Which organisation or organisations do you think are conducting this survey?" was coded by the enumerator as including any national/local NGO, by name or described as a category (enumerators could code as many options as they deemed applicable); $= 0$ otherwise

surveyorg_consultant	All except Tanzania	= 1 if respondent's unprompted, open-ended answer to "Which organisation or organisations do you think are conducting this survey?" was coded by the enumerator as including any local consultant, by name or described as a category (enumerators could code as many options as they deemed applicable), = 0 otherwise describe