
CENTRING GENDER AND POWER IN EVALUATION AND RESEARCH

Sharing experiences from Oxfam GB's quantitative impact evaluations

Sarah Barakat, Alexia Pretari and Jaynie Vonk

Bringing a feminist intent to research, monitoring and evaluation practices leads to defining these as tools to contribute to transforming the lives of women, girls and non-binary people, and to bringing about social justice. This has meant putting gender and power at the centre of our practice, which has in turn shaped the technical choices made specifically in quantitative impact evaluations. This paper focuses on describing how these technical choices, as well as ethical considerations, are changed by this feminist intent. The paper also presents the lessons learned and questions raised along the way, which may be useful for MEAL and research practitioners, as well as programme managers. How can we bring intersectionality to the fore? What does it mean to go beyond the gender binary? How can this work be transformative?

Oxfam Discussion Papers

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the authors and not necessarily those of Oxfam.

For more information, or to comment on this paper, email apretari1@oxfam.org.uk

CONTENTS

<i>Summary</i>	3
Feminist intent: what difference does it make?	3
Sampling strategies for representation and visibility	4
Choices made in measurement tools: whose experiences are represented	4
Shedding light on gender differences	4
Learning considerations: what should be done differently?	5
<i>1 Introduction</i>	7
Who is this paper for?	7
Who are we?	7
<i>2 Centring gender and power: framing</i>	9
2.1 What do we mean by gender?	9
2.2 Why do gendered impact evaluations?	9
2.3 Categorizing gender	10
2.4 Which questions can gendered impact evaluations help answer?	10
<i>3 Centring gender and power: technical considerations</i>	12
3.1 Who is listened to?	12
3.2 What matters for measurement tools?	14
3.3 Systematically looking at gender differences in individual-level outcomes	15
<i>4 Learning considerations: what should be done differently?</i>	17
4.1 Bringing intersectionality to the fore	17
4.2 Going beyond the binary	18
4.3 A gender-transformative approach?	19
<i>Conclusion: shifting power dynamics in knowledge production</i>	21
Acknowledgements	23
References	24
Notes	26

SUMMARY

This paper shares Oxfam GB's experience in bringing a feminist intent to quantitative research methodologies and centring gender and power in impact evaluations. It aims to provide practical guidance on methodologies and tools, and to provoke further reflections and changes in practices. It describes technical choices and ethical considerations, and presents lessons learned and questions raised along the way. While the paper shares experiences from impact evaluations, many aspects are more broadly relevant and will be of interest to other monitoring, evaluation, accountability and learning (MEAL) and research practitioners, as well as to programme managers.

The practices, guidance and lessons shared in this paper stem from work situated within the specific methodological and institutional frame of quantitative tools that are reductionist by design. They also reflect the experience of a large Northern international organization supporting implementation of interventions in the Global South. Bringing a feminist intent to this work leads to making practical adjustments to ensure increased representation and safety, and sets a longer-term transformative pathway to question this institutional frame itself and consider the structural changes that need to take place, some of which we highlight in the conclusion.

FEMINIST INTENT: WHAT DIFFERENCE DOES IT MAKE?

Bringing a feminist intent into research, monitoring and evaluation defines these practices as tools that bring about social justice and contribute to transforming the lives of women, girls and non-binary people. We understand social justice from an intersectional perspective: the dismantling of all systems of oppression and domination, in particular racist, capitalist, heteropatriarchal and colonialist structures, to bring about a just society. Impact evaluations have the potential to examine and assess changes towards intersectional gender equality and women's rights, and to explore differentiated impacts between women, men and non-binary people in all their diversity. By that, we mean taking into account the intersection of gender and race, class, trans history and other critical markers of structural inequalities.

Putting gender and power at the centre leads to questioning whose voices are heard, whose experiences are reflected in the data, and whose worldviews are reinforced. In the absence of explicit intent to challenge our own worldviews as practitioners, operating under the guise of so-called neutrality, or 'objectivity', can lead to strengthening the status quo and reproducing dominant narratives – sexist, racist and classist ones in particular. This paper details three technical considerations, and related ethical ones:

1. Paying attention to who is represented in the data – in other words, sampling strategies;
2. Deciding which topics are prioritized in measurement tools;
3. Systematically highlighting gender differences and differential impacts in the analysis.

While we recognize that there is more to do to bring intersectionality to the fore – including highlighting different experiences of structural inequality and oppression, as well as expanding how we can safely ask about gender non-binary and trans experiences in our evaluation practice – this paper first discusses these three points with a focus on gender as understood in a binary manner. The paper then builds and expands on this in the learning considerations section.

SAMPLING STRATEGIES FOR REPRESENTATION AND VISIBILITY

Sampling strategies play a key role in representation and visibility of different social groups, and make gendered statistical analyses possible. We want to be able to hear from women, men and non-binary people, including those who are not necessarily in positions of power in the household. The paper presents scenarios for sampling strategies, depending on whether or not a comprehensive list of individuals is available or can be gathered. If there is no such list, then a clear sampling protocol needs to be used during data gathering. We outline two options:

- Option 1: interview multiple people per household.
- Option 2: interview one person within each household, randomly varying the gender of the person to be interviewed. This paper provides guidelines and different ways to do so, including in settings where it is preferable to match the gender of interviewees and interviewers (particularly when sensitive or difficult topics are mentioned, and depending on gender norms).

Depending on context, it is important to consider interrelated aspects of identity such as race, ethnicity, caste and class. These can also influence the interviewer–interviewee relationship and the comfort and safety of interviewees; it may also bias the data gathered. These aspects must be accounted for in the composition of the interview team.

CHOICES MADE IN MEASUREMENT TOOLS: WHOSE EXPERIENCES ARE REPRESENTED

Putting gender at the centre leads our attention to power imbalances and hierarchies. For individual-level outcomes or indicators, this means including and prioritizing questions around intra-household dynamics, such as access to, and control over, information or resources, decision making and task distribution, including the distribution of unpaid care and domestic work. It also means exploring aspects of agency and confidence at the personal level, as well as experiences of violence, which are also gendered.

Given the sensitivity of some of these topics, safety considerations are essential. Ensuring privacy in the interview process and creating a safe environment is a critical role of the interviewer. Survey protocols also play a role, for example by only asking sensitive questions to one household member within each household and putting in place mechanisms to ensure that the person is alone. Additionally, developing privacy, safeguarding and protection reporting protocols is critical, as is ensuring they form part of the interviewer training. Considering risk related to the way we ask about sensitive topics has led to a shift in how we measure the prevalence of violence against women and girls in our impact evaluations, leading us to use list randomization rather than direct questions. Integrating narrative-based approaches into quantitative surveys has also brought emergent issues to the fore, allowing greater insights into gendered lived experiences.

SHEDDING LIGHT ON GENDER DIFFERENCES

Gendered impact evaluations allow exploration of differential impacts on a given individual-level outcome (behaviour, opinion, experience, etc.) with a gender lens. They make it possible to respond to the following questions: do women and men have different behaviours, opinions, experiences,

etc.? Is the intervention making a difference in their lives? There are different ways to answer these questions using quantitative methodologies.

Depending on the intervention being evaluated, the analysis will lead to measuring different types of impact (direct or indirect) and could also give insight into who participates in interventions. The approach can help in exploring the role of interrelated social characteristics (race, ethnicity, caste, age, etc.), thus enabling documentation of intersectional gender dynamics, differences and discriminations.

LEARNING CONSIDERATIONS: WHAT SHOULD BE DONE DIFFERENTLY?

When reflecting on our own practice, questions arose around how to expand the integration of a power lens into MEAL, evaluation and research more broadly for gender-transformative intent. Below, we share some of the key questions, reflections and practical considerations from this process.

How to bring intersectionality to the fore

More needs to be done to put the variety of lived experiences of women, men and non-binary people in different settings to the fore of evaluation and research. Race, ethnicity, class, sexual orientation and other dimensions will shape their experiences based on structural inequalities, and may influence the way they interact with and benefit from an intervention. Identifying these dimensions through power analysis early in the process allows for changes in the evaluation design. Building on the technical and ethical considerations developed above, the paper expands on them to discuss: 1. how sampling strategies can be adapted to be intersectional – depending on whether some aspects of identities are likely to vary within or across households; 2. how the prioritization of topics covered in survey tools will reflect the experiences of different social groups in positions of marginalization; and 3. how to carry out intersectional analyses. The paper also highlights the limitations of quantitative approaches.

How to reflect the specific experiences of trans and non-binary people

Impact evaluations we have conducted to date have mainly focused on a binary categorization of gender. Expanding gender categories in survey practice can be done in different ways that we outline in this paper. However, there is no one, single solution; choosing which approach to use requires consideration of safety in each specific situation and asking whether the data needs to be collected (is there a clear plan to use it?). The context is critical, considering that trans identities may be stigmatized and people may be subject to violence, and the act of asking people to divulge their identity may carry a risk in itself. If it is safe to ask about trans identities, protocols can be adapted accordingly. When it comes to sampling strategies, a specific (statistical) challenge arises related to the fact that trans people may represent a small share of a given population. The paper outlines a few considerations about this.

How to ensure evaluation and research is gender transformative

While challenging the worldviews reproduced in our research, evaluation and MEAL practices is critical to bring about social justice, transformation requires more. The current practices in Oxfam GB's impact evaluations take a step in the right direction for achieving gender transformation and

dismantling patriarchal systems of oppression. However, two key challenges prevent them from being gender transformative. First, the interventions being assessed sometimes include limited intersectional gender integration in their design or implementation, making analysis of transformation premature. Second, for impact evaluations to be truly transformative, they need to go beyond methodological adjustments and drive action for change and intersectional gender justice, which requires putting this wealth of valuable evidence into use. For evaluation and research to be tools towards transformation means making use of the findings; it requires activism alongside accountability.

SHIFTING POWER DYNAMICS IN KNOWLEDGE PRODUCTION

Finally, feminist intent brings attention to knowledge hierarchies and creates awareness of the power dynamics in knowledge production processes. Research and evaluation outputs represent the work of many people, and a lot of contributors are made invisible in those outputs. In the context of international development, the work of individuals from racialized groups¹ is often rendered invisible. This colonial legacy needs to be acknowledged and acted upon. Structural changes are needed to address the power imbalance; this goes far beyond small tweaks, such as improving visibility in publication. Also, feminist intent brings attention to methodological and epistemological hierarchies. A question remains about how epistemic affiliation and methodological choices limit the contribution of research and evaluation to the social justice agenda.

A note on language: in this paper, we refer to women, men and non-binary people in all their diversity. This means we take into account the intersection of gender with race, class, trans history and other critical markers and drivers of structural inequalities. In saying women, we include trans women and in saying men, we include trans men. Where we are seeking to specifically highlight the experiences of a specific population group, we make it explicitly clear.

1 INTRODUCTION

Gathering and analysing data, along with the process of designing tools to do so, comes with the potential and power to shed light on experiences of marginalization, exclusion and oppression, as well as leadership and resilience. It also comes with the risk of erasure – of making hidden factors and experiences even more invisible. For this reason, questions of representation lie at the centre of such processes and are core to research and evaluation; acknowledging these questions opens up opportunities. For example, impact evaluations can become tools to shed light on discrimination, to reflect a variety of lived experiences and material conditions, and to question whether different social groups are benefiting from an intervention differently. Bringing a feminist intent to research, monitoring and evaluation practices leads to defining these as tools to contribute to transforming the lives of women, girls and non-binary people, and to bringing about social justice. We understand social justice from an intersectional perspective: the dismantling of all systems of oppression and domination – in particular racist, capitalist, heteropatriarchal and colonialist structures – to bring about a just society.

This feminist intent has led us to put gender and power at the centre of our work, which in turn has led to making different ‘technical’ choices for impact evaluations. We draw on our experience working on/with quantitative impact evaluations, which are conducted for accountability and learning in Oxfam GB (OGB). Quantitative tools are reductionist by design, and conducting impact evaluations from and for a large Northern international NGO about projects implemented in the Global South comes with a specific position. Within these methodological and institutional constraints, and considering this position, what changes when we bring a feminist intent? This paper focuses on describing how technical choices and ethical considerations are changed by this feminist intent, as well as presenting the lessons learned and questions raised along the way.

We will first reflect on the intention and reasons for putting gender and power at the centre of impact evaluation practice. Second, we will go through some of the key technical considerations that arose from this centring, learning from our impact evaluation experience. We believe that these are applicable to other MEAL, evaluation and research practices. Finally, this process raised questions and revealed areas that need to be better integrated into research and MEAL practices. How can we bring intersectionality to the fore? What does it mean to go beyond the gender binary? Is this work transformative? We will discuss these issues here and start drawing up some potential ways forward.

WHO IS THIS PAPER FOR?

This paper shares experiences from OGB’s quantitative impact evaluations and steps taken to put gender and power at their centre. It provides practical guidance on changes in methodologies and tools and aims to provoke further reflections and changes in practices. While quantitative impact evaluations draw on specific methodologies and understanding of causality, these are not the focus of the paper. Many aspects of the approaches taken and lessons learned are relevant for other types of work. As such, this paper will be of interest to evaluation, MEAL and other research practitioners more broadly. It will also be of interest to programme managers, as it touches on aspects related to programme design and resourcing.

WHO ARE WE?

We work with OGB as Impact Evaluation and Gender-Transformative Programming leads with a global remit. This work comes from a motivation to bridge the gap between our personal feminist values

and our professional practices, including in spaces that are known for being technical and associated with a certain idea of neutrality or 'objectivity'. We recognize that our professional roles and our own positionality come with power and privileges, as do the thought traditions from which quantitative impact evaluations emerged. Integrating feminist values in impact evaluations brings attention to these power dynamics, and to the ones at play in knowledge production more broadly. In Section 2 we reflect on the methodological and institutional frame from which we draw experiences, while in the conclusion we broaden the discussion and highlight some of the structural changes that need to take place.

2 CENTRING GENDER AND POWER: FRAMING

Impact evaluations have the potential to examine and assess changes towards gender equality and women's rights, and to explore differentiated impacts between women, men and non-binary people in all their diversity (taking into account intersecting social identities and dimensions of structural inequalities, including race, ethnicity, sexuality, trans history, age, ability, etc.). As we will detail throughout this paper, 'gendered impact evaluations' differ from other impact evaluations by adapting their methodology, measurement tools, sampling and survey strategies with the intention to measure these gender justice impacts and differentiated impacts.

2.1 WHAT DO WE MEAN BY GENDER?

Before focusing on these evaluations, it is important to clarify what we mean by gender. Oxfam understands gender as a social construct that assigns roles, attributes and opportunities to women and men in a given context, and that is replicated and reinforced through norms, behaviours, laws and systems. Gender is one dimension of social and political identity and it intersects with other aspects, such as race, class, sexuality, ethnicity, caste, etc., to create and reinforce interdependent systems of marginalization, discrimination and oppression. These systems, which include heteropatriarchy, racism, colonialism, capitalism, ableism and others, all intersect to perpetrate inequality.

Oxfam's Sexual Diversity and Gender Identity policy² defines gender identity as an individual's experience of their own gender and how they wish to convey this externally and personally. This includes recognizing the experiences of transgender people, including those who identify as non-binary and all other gender identities.

Our work on gendered impact evaluations integrates both definitions, as the structural and the individual are deeply intertwined.

2.2 WHY DO GENDERED IMPACT EVALUATIONS?

Bringing feminist values and intent into impact evaluations leads to framing them as tools that contribute to transforming the lives of women and girls, as well as non-binary people, and bringing about gender and social justice. In quantitative approaches, putting gender and power at the centre leads us to question whose voices are heard, whose experiences are reflected in the data, and whose worldviews are reinforced or reproduced. Research and MEAL practices involve many trade-offs when taking into account resource constraints. The choices made are political (and who makes these choices also matters). In the absence of explicit intent to challenge our own worldviews as practitioners, operating under the guise of so-called neutrality or 'objectivity' (which is often the case with quantitative approaches) can lead to strengthening the status quo and reproducing dominant narratives – sexist, racist and classist ones in particular. This centring of gender and power dictates the technical choices and the process within the frame in which we operate.

At OGB, we conduct impact evaluations for accountability, learning and programme quality. In particular, as part of our Effectiveness Review series³ – impact evaluations conducted for a randomly selected set of mature projects – for projects reaching a large number of people, and with a primary aim to improve people's lives, we have so far used a primarily quantitative approach (see

Section 2.4). This paper reflects on evaluation practices within this body of work. Through the quantitative approaches we have been using, we have sought to maximize opportunities to shed light on and give visibility to different lived realities. In an impact evaluation, it is possible to document both discriminations and inequalities, as well as to question whether different social groups are benefiting differently from an intervention, or not. We see this documentation and questioning as key components of ensuring programme quality.⁴

2.3 CATEGORIZING GENDER

Dominant research and evaluation methods rely on categorizing gender. In quantitative approaches, this categorization takes place when data is gathered, and is then understood as a proxy for gender power dynamics.⁵ This categorization can be self-identified or identified by the researcher/interviewer. In the setting of Oxfam's impact evaluations carried out to date, often via quantitative surveys, how gender is represented in the data, alongside all gathered information, is the result of an interaction between interviewee and interviewer. Our impact evaluations have mainly used binary gender categories – i.e. women/girls and men/boys. In recent years, we have sometimes added a choice of 'other' to reflect our recognition of a variety of gender identities. We acknowledge that this is still limited and is 'othering' in itself; at the same time, allowing interviewees to self-describe is another step towards ensuring increased recognition and visibility. The choice to expand beyond binary gender categories is driven by the context (e.g. safety, language) as well as trust, the setting of the interview, and interviewer–interviewee power dynamics (discussed in detail in Section 4.2).

Additionally, while we would expect the data gathered to reflect the interviewee's reality rather than the interviewer's observation or their conception of the interviewee's gender, we cannot rule out that this may sometimes happen. Gender identity, and how this should be addressed during interviews, therefore needs to be discussed in any training prior to data collection, although training does not eliminate all possible biases.

While we seek to bring intersectionality to the fore (see Section 4.1) and expand how we can safely ask about gender non-binary and trans experiences in our evaluation practice (see Sections 4.2 and 4.3), the current approach discussed in Section 2.4 and Section 3 focuses on gender, understood in a binary manner and without differentiating between cis and trans women or cis and trans men.

2. 4 WHICH QUESTIONS CAN GENDERED IMPACT EVALUATIONS HELP ANSWER?

Integrating gender justice happens along a continuum, from interventions that do not recognize gender at all (often referred to as 'gender blind' or harmful programmes) to those that seek to sustainably and structurally dismantle systems of oppression and discrimination (what we refer to as transformative) (see Parvez Butt et al., 2019, for an example of use of the gender continuum for research planning). One of the first steps towards meaningful intersectional gender integration is to recognize differences in needs, resources, opportunities and intervention impact, and enabling meaningful disaggregation when analysing the data. An impact evaluation approach can be considered gendered when it takes gender into account through every stage of the evaluation process – evaluation design, data collection, analysis and reporting. For evaluation design and data collection, we can consider gender in sampling strategies and measurement tools. In analysis and reporting, we can systematically look for differences by gender and ensure these findings are discussed and shared.

The evaluation approach differs slightly, in terms of the type of impact that can be measured (direct or indirect) and the type of questions answered, depending on the primary focus of the intervention

being evaluated (see Table 1). For interventions at the household or community level, as well as those working directly with women and men at the individual level, gendered impact evaluations can lead to measuring the average direct effect among women and men, provided the overall sample is representative by gender. Doing so can help us understand to what extent men and women experienced impacts differently in cases where the intervention intended to benefit everyone equally, or in cases where the theory of change had a focus on benefiting households, for example. The approach described here leads to considering the diversity of individuals within the household.

However, for interventions that target one social group – for example, those that work with women only – gendered impact evaluation measures the average direct effect among the targeted group (e.g. women) and the average indirect effect among those not targeted (e.g. men). For example, this can help us understand how much women who were directly involved benefited from the intervention, as well as how much men who were not involved benefited (e.g. through intra-household sharing of knowledge and resources between women and men, or changes in social norms). It will also help us question who was targeted and/or who chose to participate in an intervention, and whether any social characteristics intersecting with gender, determined this participation (referred to as ‘targeting/selection question’ in Table 1).

Table 1: What can be measured, depending on the primary focus of the intervention

Intervention primary focus	What can gendered impact evaluations measure?
Community-level intervention and above	Average direct effect among women and men
Household-level intervention	Average direct effect among women and men
Individual-level intervention for both women and men	Average direct effect among women and men Targeting/selection question
Individual-level intervention targeting one social group, for example women only	Average direct effect among women and indirect effect among men Targeting/selection question for women

In addition to the impact and targeting/selection question highlighted in Table 1, this approach yields novel information that enables documentation of gender dynamics, differences and discriminations. The act of gathering data in the way described in this paper has a cross-cutting purpose: to make the invisible visible. This can mean documenting experiences with violence or unpaid care and domestic work, and how those experiences differ based on gender. It also means being more intentional about whose experiences are represented. Ensuring that our own methodologies and data do not hide important differences is no small feat.

3 CENTRING GENDER AND POWER: TECHNICAL CONSIDERATIONS

The decision to centre gender and power has dictated some of the technical choices in our impact evaluations. Most of these considerations are not specific to impact evaluations and would be relevant to other MEAL and research practices.

3.1 WHO IS LISTENED TO?

3.1.1 Sampling strategies and the interactions between interviewers and interviewees

Data collection can make structural inequalities invisible. Feminists – and feminist economists in particular – have been saying this for a while, yet still we have observed a tendency in data-collection processes to focus on the household level as a unit of analysis without considering intra-household dynamics. Intra-household dynamics are only one aspect, and it is critical to pay attention to who is represented in the data. From a statistical perspective, we have to be intentional in our sampling strategies to enable representation and visibility of different social groups, and make statistical analyses by social group possible. In any evaluation design and during data gathering, sampling strategies are key. They are fundamental to a gendered impact evaluation approach, but these practices can also be adapted for other MEAL or research activities. In particular, once we know which communities are included in an evaluation, we are concerned with how individuals are selected for interviews and aim for a balanced representation of women and men that is not determined by household structure. We want to be able to listen to a variety of women and men who are not necessarily in positions of power in the household, as that could lead to reinforcing patriarchal norms.

When lists of individuals or households in a specific setting are available or can be gathered

If a comprehensive list of individuals is available for sampling, the most straightforward approach is to sample randomly from that list. As long as the list includes approximately half men and half women, this should result in a gender-balanced sample. If the list also includes gender, sampling can be stratified to ensure that a representative or an even proportion of women and men are sampled (this would be particularly important if the list is not gender balanced and the proportion of one social group is really low, in a given setting).

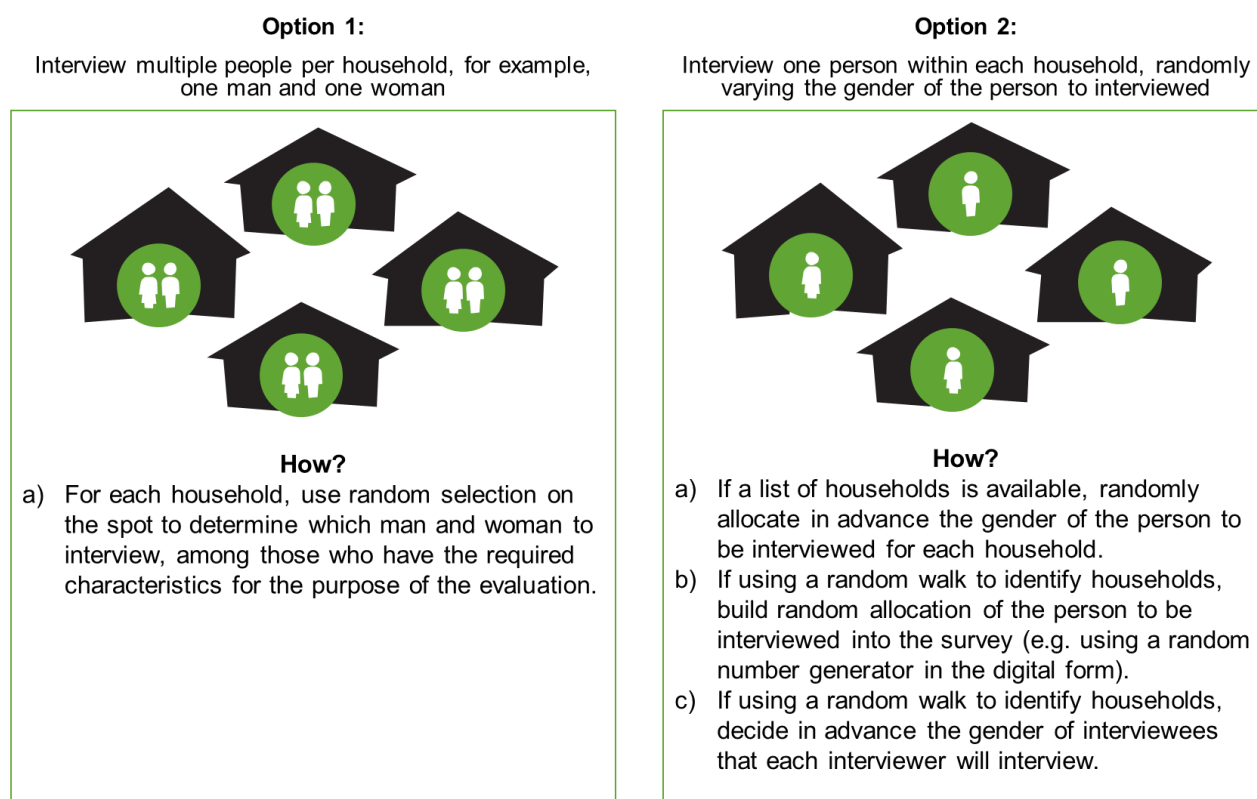
If the list available is by household rather than by individual, then a two-stage sampling strategy is needed. This means first sampling households (from the list), and then sampling individuals within each of those households (options 1a and 2a in Figure 3.1). The main limitation of this approach is that not all individuals in a given community have the same likelihood of being sampled, since the number of household members varies. However, if the number of household members is known for each household, this could be adjusted for during sampling.

When lists of individuals or households in a specific setting are not available and cannot be gathered

If there is no list of individuals or households available, and it is not possible to gather this data before sampling (e.g. due to resource limitations), then a clear sampling protocol needs to be used during data gathering. Two main options are outlined below, alongside relevant considerations (options 2b and 2c in Figure 3.1). Both of these options aim to achieve gender balance among evaluation participants while minimizing biases that could arise due to the gendered nature of availability.

First, it is important to define any characteristics required of the interviewee(s) for the purpose of the evaluation. These could include things like age range (e.g. 18 years and older), being knowledgeable about a certain topic (e.g. household water access and use), being involved in a specific activity (e.g. farming or trading) within the household, etc. Once these characteristics are defined, gendered sampling could be done within these constraints in the ways described in Figure 3.1.⁶

Figure 3.1: Options for gendered sampling, depending on availability of household-level lists



Note that for option 1 above, if several individuals within each household are to be interviewed, special considerations need to be made depending on the interview topics to be covered. For instance, sensitive topics such as domestic violence should only be discussed with one member of the household to ensure safety (see Section 3.2 for more detail on this point).⁷

In some contexts, it is necessary for women to interview women and men to interview men to create a safe environment, particularly when sensitive or difficult topics are shared. In other cases, it is completely acceptable for men to interview women and women to interview men. While this section specifically focuses on gender, it is important to consider for each evaluation, and depending on its context, which other aspects of interviewers' identity, such as race, ethnicity, caste or class, might also influence the interviewer-interviewee relationship and may thus affect the data generated.

- Taking these gender norms and power dynamics into account will have consequences for the composition of the interview team. As mentioned, it can be necessary for women to

interview women and men to interview men. For example, when aiming to have a sample with an equal split of women and men, then the team of interviewers will also be equally split between women and men. Beyond considering gender balance in the team composition, a gendered approach also examines the specific barriers or obstacles interviewers may face, such as mobility and safety issues, based on their gender, race, ethnicity, religion, etc. If interviewer and interviewee gender do not need to be matched, all the options presented in Figure 3.1 are possible.

- If interviewer and interviewee gender need to be matched in order to create a safe space:
 - Option 1 can be done by interviewer teams consisting of one woman and one man working together to conduct interviews with each household.
 - Option 2a can be done by assigning women interviewers to the households where a woman needs to be interviewed and men interviewers to the households where a man needs to be interviewed.
 - Option 2c can be done by asking women interviewers to interview a woman for each sampled household and men interviewers to interview a man for each sampled household.

Taking into account the need to ensure safety and availability of interviewees, a clear protocol needs to be agreed upon in advance for what to do when a household does not have a person of the gender to be interviewed (while meeting all other required characteristics), as well as how many times to return to a household in case of unavailability of the potential interviewee.⁸ This will also depend on whether it is deemed necessary to match interviewer and interviewee gender.

3.2 WHAT MATTERS FOR MEASUREMENT TOOLS?

Putting gender at the centre makes us attentive to power imbalances and hierarchies and requires us to intentionally explore factors of power differentials and marginalization. It means being attentive to (and making sure that measurement tools reflect) women and girls' lived experiences, and more broadly gendered experiences, and how these differ based on intersecting identity categories of race, ethnicity, caste, etc. Putting gender at the centre also leads to us asking these questions in a sensitive and safe manner. Among other features, the platform Data2X flags the need for data to 'reflect gender issues', to be 'based on concepts and definitions that adequately reflect the diversity of women and men and capture all aspects of their lives' and to be 'developed through collection methods that take into account stereotypes and social and cultural factors that may induce gender bias in the data'.⁹

Designing data-gathering tools requires us to make choices on the topics covered, taking into account the time people give for the interviews and the resource constraints. For example, we tend to aim for interviews to last 45 minutes to one hour maximum. For individual-level outcomes or indicators, centring power and gender relations will lead to prioritizing questions around access to, and control of, information or resources within the household, and decision making and task distribution within the household, including unpaid care and domestic work, for example. The Household Care Survey Toolkit (Rost et al., 2020) provides different ways of measuring aspects of unpaid care and domestic work,¹⁰ and parts of it could be used to ensure the integration of questions on this area. This prioritization will also lead to exploring aspects of agency and confidence at the personal level, as well as experiences of violence, which is gendered. It can also lead to probing on individuals' experiences in broader community and society, such as their involvement in and influence over decisions taken in community groups and awareness of legislative frameworks and knowledge of rights. The 'how to' guide to measuring women's empowerment (Lombardini, et al., 2017) provides examples of indicators.¹¹

Putting gender at the centre also requires us to ask questions in a sensitive and safe manner. This is about who is asking the questions, in what setting, and how a given question is asked. While the identity of the interviewer is discussed in Section 3.1, ensuring privacy in the interview process is key, and creating a private and safe environment is a critical part of the interviewer's role. This can be reflected in survey protocols, for example by only posing sensitive questions to one household member within each household.¹² To ensure this person is alone, one approach includes introducing a safe word;¹³ this is especially important during phone interviews because the interviewer cannot see nonverbal cues or changes in the interviewees' environment. It is also about developing privacy and safeguarding and protection reporting protocols – even when the interview is not probing about violence, such experiences can be shared – and ensuring that these core topics are discussed with interviewers during training.¹⁴

Considering risk related to the way a given question is asked has led to a shift in how we measure the prevalence of violence against women and girls in our Effectiveness Reviews.¹⁵ Following a **blog post**¹⁶ by Amber Peterman and Tia Palermo on using list randomization to measure taboo topics, including gender-based violence, we now use this methodology to assess average prevalence of violence. List randomization is a method by which people being interviewed are presented with a list of events and asked to indicate the number of events they have experienced. The list of events every interviewee is presented with is randomly allocated between two sets: one set that includes a sensitive event (an experience of violence) and one set that does not. Prevalence of violence is then calculated by comparing answers between the two sets. This choice comes with a trade-off: while it does not provide data on individual experience of violence, it does allow to estimate the average prevalence, which is what is needed from an evaluation and policy perspective. This measurement approach is much safer than the individual-level approach we had previously been using (building on the Demographic and Health Survey domestic violence module¹⁷).

Finally, integrating narrative-based approaches in a quantitative survey is one way to open the door to emergent issues. This would mean including different ways to prompt interviewees to share their experiences if they want to, using prompts inspired by Most Significant Change or SenseMaker®, or asking specific open-ended questions. Combined with attention to who is represented in the data and to analytics (see Section 3.3), narrative-based approaches can provide greater insights into gendered lived experiences (e.g. see Pretari, 2019;¹⁸ Vonk, 2021(b)¹⁹).

3.3 SYSTEMATICALLY LOOKING AT GENDER DIFFERENCES IN INDIVIDUAL-LEVEL OUTCOMES

At the analysis stage, we are able to systematically explore different questions on a given individual-level outcome (behaviour, opinion, experience, etc.) with a gender lens. First, do women and men have different behaviours, opinions, experiences, etc.? Second, is the project making a difference to their lives? This can be further broken down when estimating the average impact among women and among men, and testing whether the impact is different between women and men (are there differential impacts for different social groups?).

There are different ways to answer these questions in a quantitative setting, depending on which questions are prioritized and on the impact evaluation design (experimental, difference-in-differences, matching, etc.) for questions related to impact. For example, in the Effectiveness Review series, we often use a statistical technique called propensity score matching (PSM) to estimate the average impact among all people being interviewed. In addition, to estimate gender differences, impact among men, impact among women and to test for the significance of the difference between these impacts, we run propensity score (PS)-weighted regressions with interaction terms (e.g. Pretari, 2019; Pretari, 2021; Vonk, 2021a; Vonk, 2021b; Vonk, 2021c).^{20,21} This leads to systematically presenting and commenting on these results in the evaluation products, and sparks discussions from which to draw learning recommendations.

While we build on our experience in impact evaluations, this is also relevant to other evaluations or research projects. If gender difference is the focus, descriptive statistics with systematic gender-disaggregation will enable us to shed light on gender differences, depending on the sampling choices. The representation of gender-sensitive and power-aware theories of change can also be great analytical tools, alongside statistical analysis.

4 LEARNING CONSIDERATIONS: WHAT SHOULD BE DONE DIFFERENTLY?

When reflecting on our own practice, questions came up around how to expand the integration of a power lens more broadly into MEAL and other research for gender-transformative intent. In this section we share some of the key questions and reflections from this process. In doing so, we start to identify areas for development and further change in practice and systems.

4.1 BRINGING INTERSECTIONALITY TO THE FORE

Focusing on lived experiences and material living conditions of women and men is key to unveiling structural inequalities but carries a risk of essentialization (Sigle-Rushton, 2014), particularly if the causes of any observed differences are left unexplored (see Jenny Chanfreau's blog post²²). We need to go beyond the points outlined in Section 3 to acknowledge the variety of lived experiences that women in a given setting may face, and that women and men are not homogenous categories. Race, ethnicity, class, sexual orientation and other dimensions of structural inequalities will shape their experiences, and may shape the way they interact with an intervention and benefit from it.

Intersectionality is a concept that is rooted in Black feminist activism and theory.²³ It requires us to interrogate the different systems of power and related social hierarchies such as gender, race and class in particular. Bringing intersectional frameworks to impact evaluations thus requires us to identify the different dimensions that shape structural inequalities in a given context, and the lives of women and girls in particular, and to unpack the complexity of power relations. A key question to ask first is: how safe is it for the people who will disclose information, and for the interviewer to record this information through an interview process?

Identifying these dimensions through power analysis early in the process allows for changes to be made in the evaluation design. We will start outlining the ways in which the considerations above could lead to the evaluation design being adjusted accordingly. While the technical considerations laid down in Section 3 will still be core aspects of the evaluation or research design, bringing in an intersectional lens will affect some of the choices made within each point.

4.1.1 Adapting sampling strategies

For some dimensions of structural inequalities, the sampling strategies described in Section 3.1 will apply as they can be easily tweaked.

Race or ethnicity may vary within households in different settings; for example, this variation is often less where inter-racial or inter-ethnic marriages are stigmatized. In this case, the variation will mainly be across households. The sampling approaches described in Section 3.1 will enable representation of women and men of different ethnic groups based on the characteristics of the population, as long as the overall sample frame is representative. In cases where the share of a particular group of interest is small (for example Hindu women in a majority Muslim society, or black women in a white supremacist society), the data generated by representative sampling will bring about a small number of interviews with this group – most likely too small to enable meaningful statistical analysis. The protocols will therefore have to change further (discussed in Section 4.2).

Some dimensions will vary within households. There is likely to be a spread of ages of women and men in any given household, for example. The protocols described in Section 3.1 could easily vary age and gender together to examine impact at the intersection of the two factors. The sampling

protocol could randomly identify who to interview within the household: for example, a woman below 30, a man below 30, a woman above 30, a man above 30. The details of the protocol will have to be worked out (e.g. who should be the interviewer? What to do when there is no one from this gender and age group in the household? If the person is not available at the time of the interview, how many times will the interviewer come back? What to do if another household member wants to answer the survey?).

4.1.2 Prioritizing areas covered in measurement tools

The areas covered in measurement tools will have to be prioritized to cover specific discriminations faced by specific social groups in positions of marginalization, or areas that are key to their lived experiences. For example, access to education may be particularly restricted for young girls of a certain ethnic identity in a given context, owing to social norms, availability of public services, and/or discriminatory laws and policies. This could lead to the inclusion of questions related to the issue.

4.1.3 Carrying out intersectional data analysis

At the analysis stage, several options are possible. First, research and evaluation that are not exploring impact questions can use descriptive statistics with systematic disaggregation based on the dimensions identified in the power analysis (and considered in the sampling approach). Second, similar to the approach to date in OGB's quantitative impact evaluations, regression models with triple interactions would allow examination of intervention impact by sub-groups and testing for the differences, combining two characteristics (gender and age, for example). It would also allow us to look at how such social groups experience things differently in the absence of the intervention. It does raise a conceptual question as regressions assume additivity, while intersectionality does not. Finally, the more aspects of identity and intersections we consider, the harder the analysis gets from an econometric perspective, and separated regression models per sub-group may be needed, with tests across models.

Two overall limitations must be highlighted in a quantitative framework. First, to be able to meaningfully interpret results related to multiple social groups, we need to have information on enough people in any given group, and a large enough sample size overall. Second, while quantitative approaches enable description and the shedding of light on some of the systemic discriminations faced and potential differential impacts of interventions, they place less attention on the uniqueness of experiences shaped by the multitude (and sometimes changing nature) of identities an individual goes through life with. Qualitative methods are better suited to shedding light on the nuance and complexity of identities and experiences.

4.2 GOING BEYOND THE BINARY

The platform Data2X²⁴ highlights that, 'We additionally recognize the need for data systems to evolve further to adequately and appropriately represent individuals of all gender identities, whether they identify as men, women, or other.' As mentioned in Section 2.4, the impact evaluations we have conducted to date have mainly focused on a binary categorization of gender. Expanding gender categories in survey practice can be done in different ways. Edge Effect and 42 degrees highlight two different ways – relying on one survey question or two – and discuss the pros, cons and nuances of each in a note around collecting and analysing data about people of diverse Sexual Orientations, Gender Identities and Expressions and Sex Characteristics. The Positive Voices survey developed by Public Health England (and presented in D'Ignazio and Klein, 2020) provides another way, relying on two survey questions.²⁵

There is no blanket solution to extending gender categories, and ultimately it will depend on:

1. Use – there is no need to gather information if it will not be used.
2. Context – this should be considered to decide which approach is more appropriate and culturally sensitive, and to choose the language that will reflect the way trans people want to be represented.
3. Context again – to decide whether one method is safer than the other, and whether it is safe to gather any aspect of the information.

As mentioned above, protocols can be adapted to go beyond a binary analysis. Extending gender categories in the survey will lead to intersectional analyses with similar implications when it comes to the choices made in the measurement tools and in the analysis as described in Sections 4.1.2 and 4.1.3.

When it comes to sampling strategies, a specific (statistical) challenge arises related to the fact that people who identify as trans may represent a small share of a given population. The challenge lies in the fact that the impact evaluations we conduct do not have the resources for conducting censuses prior to sampling, and rely on on-the-spot sampling.²⁶ The protocols discussed in Section 3.1 can be adapted, however, to ensure representation of trans and non-binary people (e.g. by systematically interviewing trans and non-binary people in given households – purposive sampling). Additional calculations and/or use of sampling weights may be needed afterwards. These sampling challenges are not unique to the representation of different gender identities; indeed, they are also applicable to the representation of other intersecting identity categories that may represent a small share of a given population, as covered in Section 4.1 on intersectionality.

To expand further on safety, the context will be critical, considering that trans and non-binary identities may be stigmatized and people may be subject to violence, and that the act of gathering data and asking people to divulge their identity may carry a risk in itself. Issues of wellbeing and safety need to be prioritised when thinking through whether this data should be gathered at all in any given context. The opportunity to highlight how the intersection of systems of oppression shape experiences, and how different people benefit from Oxfam’s interventions, should not come at the cost of putting people at risk. This would perhaps be more the case if the interventions were blind to gender. Would it be ethical for an impact evaluation to shed light on trans/non-binary experiences while the intervention it is assessing does not put their experiences at its core in terms of project inclusion or of systemic change?

At the same time, and if the data can be gathered safely, evaluation and research provide the opportunity to shed light on the lived experiences of different people. It is thus also important to ask whether documentation and visibility could be an aim in itself, understood as a means to bring about changes in programme design and channel resources through donor advocacy. We see visibility and representation as a core intent of quantitative methods, but marginalized groups who are sometimes already visible, and for whom this visibility can be a source of daily struggle and discrimination, may not want the extra attention. Working with LGBTQIA+ activists when it comes to visibility of trans experiences – and social justice activists more broadly, depending on the issues considered – to assess the need for data as a potential tool for advocacy, could help mitigate this.

4.3 A GENDER-TRANSFORMATIVE APPROACH?

Going back to the gender justice continuum we referred to in Section 2.4, gender-transformative research is expected to analyse and build an evidence base for longer-term, sustainable and structural shifts towards gender justice. It aims to dismantle patriarchal systems of oppression, not just work within them.

The current practice in impact evaluations at Oxfam to consider gender in evaluation design, data collection, analysis and reporting (as per the above sections) is a step in the right direction, as are reflections and practices around gender categories and intersectionality. Gender transformation is also driven through a deeper reflection – and inclusion within protocols – of considerations of

power dynamics and their relationship to violence. Recognition of the power imbalance between interviewer and interviewee, the inclusion of safeguarding, physical sexual exploitation and abuse (PSEA) in the interviewer training pack, and the development of disclosure reporting procedures also contribute to a feminist perspective that centres on, explores and attempts to transform power. Moves towards exploring and documenting intra-household power dynamics contribute to the same goal. Assessing exposure to violence, including using the recently trialled list randomization methodology as mentioned in Section 3.2, also constitutes an integral part of a gender-transformative approach.

Two key challenges remain. First, while impact evaluations have made strides towards gender transformation, the interventions they assess sometimes only include limited gender integration in their design or implementation, making analysis of gender transformation premature or superficial, and keeping the evaluation towards the lower end of the transformation spectrum. The same can be said, perhaps even to a greater extent, of intersectional gender transformation, as we still have some way to go in designing and implementing interventions that seek to dismantle intersectional systems of oppression, such as racism and colonialism. Second, for impact evaluations to be truly transformative, they need to go beyond methodological adjustments and drive action for change and intersectional gender justice. The wealth of valuable evidence they produce (including by collecting data on non-intervention-specific variables) must be put to use in reshaping strategies and interventions, designing and implementing new ones, allocating financial and technical resources, and/or exploring partnerships. For example, the gendered approach adopted in the Effectiveness Review of the Enterprise Development Programme (EDP)²⁷ in Rwanda showed gender-differentiated impact in its results, highlighting some of the structural constraints that women face in general, and depending on whether they were living with a partner/husband or leading the family in particular. This came at a time where the EDP programme was starting to develop its new strategy. The timeline alignment enabled the results to contribute to developing a strategy with women's economic justice at its core.

CONCLUSION: SHIFTING POWER DYNAMICS IN KNOWLEDGE PRODUCTION

Bringing a feminist intent into research and evaluation leads to framing these as tools for social justice. The intent dictates the 'technical' choices within the methodological, epistemological and institutional frame in which these choices are made. This frame represents an important set of constraints and limits the intent itself, and we will come back to it.

The starting point of our journey has been binary, with a focus on systematic gender sensitivity and consistent integration of a gender lens. The journey is evolving towards bringing intersectionality to the fore – understood in a political manner, i.e. with attention to systems of power – and being trans and non-binary inclusive. Ultimately it is all about the context: understanding systems of power and how they intersect in a specific context to shape individual experiences of oppression and domination, and how they shape risks and safety while data-gathering activities are carried out. Bringing intersectionality to the fore leads us to question the 'ready-made' category that the sector, including Oxfam, has relied on. For example, what is meant by headship? What are we actually examining when an analysis focuses on 'female-headed' households? What does the use of this concept say about our understanding of power and how it is shaped and exercised within and across households? Does its use reproduce a patriarchal type of power? What does it say about our assumptions of what constitutes a household, and where have these assumptions come from?

While challenging the worldviews reproduced in our research and evaluation practices is critical to bring about social justice, transformation requires more. For research and evaluation, to be tools towards transformation means making use of their findings; it requires activism alongside accountability.

Finally, feminist intent leads to attention to knowledge hierarchies. First, it creates awareness of the power dynamics in knowledge production processes and output knowledge products. For example, research and evaluation is the work of many people, yet many contributors are made invisible in the output product. In the context of international development, the work of individuals from racialized groups is often rendered invisible. We have taken steps to make this work visible in our evaluation products. But we need to go beyond small tweaks, as addressing the issue goes far beyond visibility in publication: structural changes are required to address the power imbalance. Eliséé Cirhuza highlights the 'colonizer-colonized' legacy in academia and calls for transparent and collaborative approaches in their blog as part of the (Silent) Voices Bukavu series.²⁸ In her post on the American Evaluation Association blog,²⁹ Vidhiya Shanker calls for structural changes to dismantle the systemic erasure of the work of women of colour and indigenous women in evaluation: in terms of flow of resources, i.e. money and time, but also citations, following Sara Ahmed.³⁰ Second, feminist intent leads to attention being paid to methodological and epistemological hierarchies. How do epistemology and choices of methodology themselves limit the contribution of research and evaluation to the social justice agenda?

To go further and for inspiration:

1. Data Feminism, by Catherine D'Ignazio and Lauren F. Klein; book: <https://data-feminism.mitpress.mit.edu/> and videos: <https://datafeminism.io/blog/book/data-feminism-reading-group/>. And in particular:
 - Chapter 4 'What gets counted counts', about rethinking binaries and hierarchies;
 - Chapter 5 'Unicorns, Janitors, Ninjas, Wizards, and Rock Stars', about embracing pluralism (feminist ways of knowing in data work. Data for good or data for co-liberation?).
2. The (Silent) Voices Bukavu Series, a blog series which highlights 'the premeditated violence that persists in the process of academic knowledge production', arguing 'that this process is, among other things, responsible for the dehumanization and the erasure of researchers from the Global South': <https://www.gicnetwork.be/silent-voices-blog-bukavu-series-eng/>
3. Afrofeminist Data Future, by Neema Iyer, Chenai Chair and Garnett Achieng: <http://pollicy.org/feministdata/> (available in English, French and Portuguese), and in particular the risk of re-traumatization of interviewees.
4. 'Intersectionality Matters!'. A podcast series with Kimberlé Crenshaw, bringing together panels of researchers and activists to analyse different topics in the North American context with an intersectional lens: <https://soundcloud.com/intersectionality-matters>
5. For French speakers, the 'Kiffe ta race' podcast, from Rokhaya Diallo and Grace Ly, inviting researchers, artists and activists and bringing an intersectional lens to racial issues in the French context: <https://soundcloud.com/kiffe-ta-race>
6. Is household headship a useful concept? A Research and Policy conversation: <https://www.cgdev.org/event/household-headship-useful-concept-research-and-policy-conversation>. The panellists present how the concept came about in quantitative surveys as an organizing principle, and Caren Grown discusses its limitations and how to do without it.
7. For French speakers and comic enthusiasts, two graphic representations of the headship concept may be appreciated on Pénélope Bagieu's Instagram page.
 - An original by the artist sharing her personal experience: https://www.instagram.com/p/CKTj0s5M7QG/?utm_source=ig_web_copy_link
 - A reproduction of a Mafalda strip by Quino: https://www.instagram.com/p/CFxLVbUFTeu/?utm_source=ig_web_copy_link

ACKNOWLEDGEMENTS

We would like to acknowledge the incredible contribution that everyone involved in OGB's quantitative impact evaluations across the years has made to this work. In particular, we want to acknowledge evaluation participants who gave their time to take part in surveys, interviewers who carried out the interviews, and programme and MEAL staff from Oxfam and many partner organizations who took part in evaluation design processes and result interpretations.

We also would like to acknowledge the contribution of Simone Lombardini in shaping and supporting this work, and his leading role in integrating safer measurements of prevalence of domestic violence. We would like to acknowledge the contribution of Katrina Barnes, Helen Bunting, Helen Wishart and Julia Bruce in the edition and publication of this paper. Thanks are due to Mona Mehta, Rita Soares, Patricia Espinoza Revollo, Itishree Sahoo, Helen Wishart and Fenella Porter for their critical inputs reviewing this paper.

Finally, huge thanks are due to the Power Shift project, the BAME community network and the LGBTQIA+ network for the work done internally through the sharing of resources and organizing of events. This work has been critical to support the integration of an anti-racist decolonial and intersectional feminist lens to OGB's work."

REFERENCES

- Ahmed, S. (2013). Making Feminist Points. <https://feministkilljoys.com/2013/09/11/making-feminist-points/>
- Bonilla-Silva, E. (1997). Rethinking Racism: Toward a Structural Interpretation. *American sociological review*, 465–480.
- Chanfreau, J. (2017). Why feminism: On quantitative analysis and divergent understandings of gender. LSE Department of Gender Studies. <https://blogs.lse.ac.uk/gender/2017/10/10/why-feminism-on-quantitative-analysis-and-divergent-understandings-of-gender/>
- Cirhuza, E. (2020). Taken out of the picture? The researcher from the Global South and the fight against 'academic neo-colonialism'. The Bukavu Series. <https://www.gicnetwork.be/taken-out-of-the-picture-the-researcher-from-the-global-south-and-the-fight-against-academic-neo-colonialism/>
- D'Ignazio, C. & Klein, L.F. (2020). *Data Feminism*. The MIT Press. <https://data-feminism.mitpress.mit.edu/>
- edge effect, 42 degrees Library. (N.D.). Sexual Orientation, Gender Identity and Expression, and Sex Characteristics (SOGIESC) Data Collection.
- Galandini, S. & Mager, F. (2020). *Research Ethics: a Practical Guide*. Oxfam GB. <https://policy-practice.oxfam.org/resources/research-ethics-a-practical-guide-621092/> doi: 10.21201/2020.6416
- Harris, M. & Pretari, A. (2021). *Going Digital 6: Computer-Assisted Telephone Interviewing (CATI): Lessons Learned from a Pilot Study*, Oxfam GB. <https://policy-practice.oxfam.org/resources/going-digital-computer-assisted-telephone-interviewing-cati-lessons-learned-fr-621188/>
- Lombardini, S., Bowman, K. & Garwood, R. (2017). A 'how to' guide to measuring women's empowerment: Sharing experience from Oxfam's impact evaluations. Oxfam GB. <https://policy-practice.oxfam.org/resources/a-how-to-guide-to-measuring-womens-empowerment-sharing-experience-from-oxfams-i-620271/>
- Lombardini, S. & McCollum, K. (2017). Using internal evaluations to measure organizational impact: A meta-analysis of Oxfam's women's empowerment projects – Discussion Paper. Oxfam GB. <https://policy-practice.oxfam.org/resources/using-internal-evaluations-to-measure-organizational-impact-a-meta-analysis-of-620348/>
- Parvez Butt, A., Jayasinghe, N. & Zaaroura, M. (2019). Integrating gender in research planning. Oxfam GB. <https://policy-practice.oxfam.org/resources/integrating-gender-in-research-planning-620621/>
- Peterman, A. & Palermo, T. (2018). Measuring taboo topics: List randomization for research on gender-based violence. UNICEF connect Evidence for Action. <https://blogs.unicef.org/evidence-for-action/measuring-taboo-topics-list-randomization-for-research-on-gender-based-violence/>
- Pretari, A. (2019). Active Citizenship in Tanzania: Impact evaluation of the 'Governance and Accountability through Digitalization' project. Oxfam GB. <https://policy-practice.oxfam.org/resources/active-citizenship-in-tanzania-impact-evaluation-of-the-governance-and-accounta-620855/>
- Pretari, A. (2021a). Resilience in The West Bank: Impact evaluation of the 'From Emergency Food Security to Durable Livelihoods: Building Resilience in the Occupied Palestinian Territory' project. Oxfam GB. (forthcoming)
- 24 Centring Gender and Power in Evaluation and Research: Sharing experiences from Oxfam GB's quantitative impact evaluations

Pretari, A. (2021b). Livelihoods in the Eastern Province of Rwanda: Impact evaluation of the Enterprise Development Programme's support to Tuzamurane cooperative. Oxfam GB. (forthcoming)

Rost, L., Parkes, A. & Azevedo, A. (2020). Measuring and Understanding Unpaid Care and Domestic Work: Household Care Survey Toolkit. Oxfam. <https://policy-practice.oxfam.org/resources/measuring-and-understanding-unpaid-care-and-domestic-work-household-care-survey-621082/> doi: 10.21201/2020.6775

Sigle-Rushton, W. (2014). 'Essentially Quantified? Towards a More Feminist Modelling Strategy,' in Evans, M., Hemmings, C., Henry, M., Johnstone, H., Madhok, S., Plomien, S. and Wearing, S. (eds.). The SAGE Handbook of Feminist Theory. Sage Publications.

Shanker, V. (2020). The Invisible Labor of Women of Color and Indigenous Women in Evaluation. AEA 365. <https://aea365.org/blog/the-invisible-labor-of-women-of-color-and-indigenous-women-in-evaluation-by-vidhya-shanker/>

Torre, M. & Lombardini, S. (2019). Podcast: The challenges of measuring women's empowerment. REAL Geek series, Oxfam GB. <https://views-voices.oxfam.org.uk/2019/10/podcast-the-challenges-of-measuring-womens-empowerment/>

Vonk, J. (2021a). Sustainable Water and Sanitation in DRC: Impact evaluation of the 'Sustainable WASH in Fragile Contexts (SWIFT 1)' project. Oxfam GB (forthcoming).

Vonk, J. (2021b). Sustainable Water and Sanitation in Sierra Leone: Impact evaluation of the 'Improved WASH Services in WAU and WAR Districts' project. Oxfam GB (forthcoming).

Vonk, J. (2021c). Sustainable Water and Sanitation in Zambia: Impact evaluation of the 'Urban WASH' project. Oxfam GB. <https://policy-practice.oxfam.org/resources/sustainable-water-and-sanitation-in-zambia-impact-evaluation-of-the-urban-wash-621152/>

NOTES

- ¹ Oxfam does not use race as a biological category, but as a social construct. The term 'racialized groups' is used to refer to all groups that do not enjoy the privileges of White people as a result of the socially constructed process of racialization. A racialized social system is 'one where economic, political, social and ideological levels are partially structured by the placement of actors in categories or racial groups'. Some societies are highly racialized. In others stratification does not flow along racial lines, but along ethnicity within the same racial context, as in many African and Asian countries, or along caste-based lines, in countries where the caste system is the premier systemic oppression. (Bonilla-Silva, 1997).
- ² Oxfam Sexual Diversity and Gender Identity Rights Policy: https://www-cdn.oxfam.org/s3fs-public/file_attachments/story/sexual_diversity_and_gender_identity_rights_policy.pdf
- ³ Oxfam Effectiveness Review Impact Evaluations: <https://policy-practice.oxfam.org/series/effectiveness-reviews/>
- ⁴ This body of impact evaluation work takes place within Oxfam GB's Strategic Evidence Framework, which reflects an organizational commitment to integrate a gender lens into impact evaluations since 2019 (following several years of work learning and practising how to adapt our approaches).
- ⁵ Jenny Chanfreau in her blog states that, 'In the context of quantitative analysis, which lends itself well to describing and generalising patterns, I think it's not just about being clear about stating that the individual-level male/female categorical variable in the survey analysis is understood as a proxy, however imperfect, for the subtle and amorphous gender.' <https://blogs.lse.ac.uk/gender/2017/10/10/why-feminism-on-quantitative-analysis-and-divergent-understandings-of-gender/> Different questions are usually asked in survey tools to create this categorical variable. These provide a different understanding of this proxy: sex assigned at birth, civil status – for which the categorization will depend on the laws in place in a given country – or the gender an individual identifies with.
- ⁶ Intersectional sampling strategies and analysis are discussed in Section 4.1.
- ⁷ Additional considerations need to be taken for option 2 as well, for particularly sensitive questions such as direct questions related to prevalence of violence.
- ⁸ For example, see the report of the impact evaluation of the 'Sustainable WASH in fragile contexts (SWIFT 1)' project in DRC (Vonk, 2021), which presents the random walk protocol with allocation of gender of the interviewee using SurveyCTO on page 18 (option 2b without matching of gender of interviewee and interviewer), or page 15 of the report of the 'Improved WASH Services in WAU and WAR Districts' project in Sierra Leone (Vonk, 2021). The Impact Evaluation of the 'From Emergency Food Security to Durable Livelihoods: Building Resilience in the Occupied Palestinian Territory' project (Pretari, 2021) presents the protocol on pages 21 and 22 (option 2c with matching of gender of interviewer-interviewee only for women interviewees).
- ⁹ Data2x website. 'What is Gender?'. <https://data2x.org/what-is-gender-data/>
- ¹⁰ Rost, L., Parkes, A. & Azevedo, A. (2020). Measuring and Understanding Unpaid Care and Domestic Work: Household Care Survey Toolkit. Oxfam. <https://policy-practice.oxfam.org/resources/measuring-and-understanding-unpaid-care-and-domestic-work-household-care-survey-621082/>
- ¹¹ Lombardini, S., Bowman, K. & Garwood, R. (2017). A 'how to' guide to measuring women's empowerment: Sharing experience from Oxfam's impact evaluations. Oxfam GB. <https://policy-practice.oxfam.org/resources/a-how-to-guide-to-measuring-womens-empowerment-sharing-experience-from-oxfams-i-620271/>. Since the publication of this guide, we have developed vignette-based measurement for the household decision-making questions to better capture intra-household dynamics. For more on this, listen to the podcast 'The challenges of measuring women's empowerment', published on the REAL Geek series: <https://views-voices.oxfam.org.uk/2019/10/podcast-the-challenges-of-measuring-womens-empowerment/>
- ¹² For particularly sensitive questions, such as direct questions related to prevalence of violence, survey protocols need to consider the risks associated with asking the same questions across households in a given area as well.
- ¹³ A pre-agreed-upon word that the interviewee can say if they no longer feel safe continuing the interview, in which case the interviewer will switch to a neutral topic briefly before ending the interview.
- ¹⁴ For key considerations in the setting of phone interviews, see Going Digital 6: <https://policy-practice.oxfam.org/resources/going-digital-computer-assisted-telephone-interviewing-cati-lessons-learned-fr-621188/>. For more training material on Safeguarding and MEAL, see the internal guidelines from Oxfam International. For more on research ethics in general, see Research Ethics: A practical guide: <https://policy-practice.oxfam.org/resources/research-ethics-a-practical-guide-621092/>
- ¹⁵ Following the management response from Oxfam GB to the Women's Empowerment meta-analysis (Lombardini & McCollum, 2017). <https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620348/dp-meta-analysis-womens-empowerment-projects-061017-en.pdf?sequence=1&isAllowed=y>
- ¹⁶ Peterman, A. & Palermo, T. (2018). Measuring taboo topics: List randomization for research on gender-based violence. <https://blogs.unicef.org/evidence-for-action/measuring-taboo-topics-list-randomization-for-research-on-gender-based-violence/>

- ¹⁷ The Demographic and Health Surveys (DHS) Program. Domestic Violence Module. https://dhsprogram.com/pubs/pdf/DHSQMP/domestic_violence_module.pdf.pdf
- ¹⁸ The use of a SenseMaker® inspired approach leads to stories related to experience of violence against women and girls in the community being shared by local animators – see pages 44 and 45.
- ¹⁹ The use of a SenseMaker® inspired approach leads to stories related to experience of violence against women and girls in the community being shared by interviewees – see Section 6.4, Sustainable Water and Sanitation.
- ²⁰ We are focusing here on cases where one individual per household took part in a survey. In cases where several household members took part, the analysis will have to be conducted separately for each group of individuals (women and men in this case) if household-level information is used as a control variable.
- ²¹ From a statistical perspective, the multiplication of regressions leads to increasing the risk of ‘false positives’ (type II errors), identifying a statistically significant result by chance. This is the case for every evaluation using multiple regressions to test various hypotheses (impact on several outcomes, for example) and we mitigate this risk by writing a pre-analysis plan – the gender lens is core to the considered analysis – as well as by documenting it in the evaluation reports. You can find an example of the ‘risk of bias table’ used from 2018 onwards in the reports published in 2021.
- ²² Chanfreau, J. (2017). Why feminism: On quantitative analysis and divergent understandings of gender. LSE Department of Gender Studies. <https://blogs.lse.ac.uk/gender/2017/10/10/why-feminism-on-quantitative-analysis-and-divergent-understandings-of-gender/>
- ²³ Intersectionality emerged as a concept in the US from black feminist activists, with the work in the seventies of the Combahee River Collective, a Black feminist lesbian socialist organization, and the term was later coined by Kimberlé Crenshaw in 1989. For more on intersectionality, you can see Kimberlé Crenshaw’s Ted Talk: <https://www.youtube.com/watch?v=ak0e5-UsQ2o>
- ²⁴ Data2x website. ‘Important data about women and girls is incomplete or missing’. <https://data2x.org/>
- ²⁵ A common characteristic of the different methods is that allowing interviewees not to answer a specific question is critical. It can be dealt with in the analysis stage. If one realizes during the pilot or at the beginning of the data-gathering process that a large proportion of people prefer not to say, one may want to query again the risks of asking the question at all, as well as reconsidering the safety and privacy of the interview setting.
- ²⁶ It would also be the case if lists of households and individuals exist but only include a binary categorization of gender or lack information related to trans history.
- ²⁷ For more on EDP, see: <http://edp.oxfam.org.uk/>. For more on the impact evaluation, see Livelihoods in the Eastern Province of Rwanda: Impact evaluation of the Enterprise Development Programme’s support to Tuzamurane cooperative (Pretari, 2021).
- ²⁸ Cirhuza, E. (2020). Taken out of the picture? The researcher from the Global South and the fight against ‘academic neo-colonialism. The Bukavu Series. <https://www.gicnetwork.be/taken-out-of-the-picture-the-researcher-from-the-global-south-and-the-fight-against-academic-neo-colonialism/>. ‘Because it is not just a question of giving Southern researchers visibility in a publication. Their involvement beginning with the design phase of the research is essential... What’s more, the full engagement and participation of Southern researchers during the entire research project would also rectify existing imbalances of power, and could thus put an end to the “colonizer-colonized” legacy, which leaves its marks to this day. This perspective could help the Southern researcher take real and tangible ownership of the project. It could also be an answer to certain hidden questions that the research assistant has about the purpose of the project and the objectives of the project commissioner or coordinator. Such approach would represent a radical rupture with the dominant idea that whoever pays has the right to determine a person’s place in the research cycle. It would challenge the idea that it is but a favor to include the Southern researcher up to the publication phase, and propose a more transparent and collaborative approach. While research project commissioners and coordinators should consider the suggestions of research assistants, why not involve them in all phases of the research?’
- ²⁹ Shanker, V. (2020). The Invisible Labor of Women of Color and Indigenous Women in Evaluation. AEA 365. <https://aea365.org/blog/the-invisible-labor-of-women-of-color-and-indigenous-women-in-evaluation-by-vidhya-shanker/>
- ³⁰ Ahmed, S. (2013). Making Feminist Points. <https://feministkilljoys.com/2013/09/11/making-feminist-points/>. <https://feministkilljoys.com/2013/09/11/making-feminist-points/>

Oxfam Discussion Papers

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the author and not necessarily those of Oxfam.

For more information, or to comment on this paper, email apretari1@oxfam.org.uk

© Oxfam International July 2021

This publication is copyright but the text may be used free of charge for the purposes of advocacy, campaigning, education, and research, provided that the source is acknowledged in full. The copyright holder requests that all such use be registered with them for impact assessment purposes. For copying in any other circumstances, or for re-use in other publications, or for translation or adaptation, permission must be secured and a fee may be charged. E-mail policyandpractice@oxfam.org.uk.

The information in this publication is correct at the time of going to press.

Published by Oxfam GB for Oxfam International under ISBN 978-1-78748-778-9 in July 2021.
DOI: 10.21201/2021/7789

Oxfam GB, Oxfam House, John Smith Drive, Cowley, Oxford, OX4 2JY, UK.

OXFAM

Oxfam is an international confederation of 21 organizations, working with its partners and allies, reaching out to millions of people around the world. Together, we tackle inequalities to end poverty and injustice, now and in the long term – for an equal future. Please write to any of the agencies for further information or visit www.oxfam.org

Oxfam America (www.oxfamamerica.org)

Oxfam Aotearoa (www.oxfam.org.nz)

Oxfam Australia (www.oxfam.org.au)

Oxfam-in-Belgium (www.oxfamsol.be)

Oxfam Brasil (www.oxfam.org.br)

Oxfam Canada (www.oxfam.ca)

Oxfam Colombia (lac.oxfam.org/countries/colombia)

Oxfam France (www.oxfamfrance.org)

Oxfam Germany (www.oxfam.de)

Oxfam GB (www.oxfam.org.uk)

Oxfam Hong Kong (www.oxfam.org.hk)

Oxfam IBIS (Denmark) (www.oxfamibis.dk)

Oxfam India (www.oxfamindia.org)

Oxfam Intermón (Spain) (www.oxfamintermon.org)

Oxfam Ireland (www.oxfamireland.org)

Oxfam Italy (www.oxfamitalia.org)

Oxfam Mexico (www.oxfammexico.org)

Oxfam Novib (Netherlands) (www.oxfamnovib.nl)

Oxfam Québec (www.oxfam.qc.ca)

Oxfam South Africa (www.oxfam.org.za)

KEDV (www.kedv.org.tr)

