CARE Cross-project Collectives Analysis: Technical Appendix

The CARE approach to development support and women's economic empowerment is often based on the building blocks of collectives. Some of these collectives are women's collectives and some are mixed gender composition. In addition, the leadership structure of these collectives can be single gender of mixed gender. The collectives design has many potential advantages. These advantages are thought to include practical opportunities such as increased access to inputs and training and higher income for sales and more access to credit. Collectives are expected to have social benefits such as changing social norms to be more gender inclusive, and increasing participants' sense of autonomy and social capital.

Individual CARE projects have completed research to assess the strengths and weaknesses of their collectives approach. The existence of the datasets from these individual projects presents an opportunity to also address more global questions on the varying effects of collectives. By combining datasets across CARE collective projects we are able to answer questions that transcend local contexts and provide evidence and guidance for continued women's empowerment work based on the foundation of collectives.

The results described throughout this document are based on a series of data manipulations and statistical models that combine all available data from the three projects. The advantage to combining datasets from across projects include larger sample sizes which leads to reduced sampling errors and the ability to control for within person, within group and within country variability, which makes it more likely to accurately locate subtle effects. The data has been gathered from a wider variety of contexts so they are less likely to be measuring local or contextual artifacts. Within-team measurement error is also mitigated. Multiple sources of error are lowered and the results enjoy improved precision. We can avoid misleading results due to concentrating on a unique subset of a population.

Databases
For the purposes of this research, we wanted to explore the effects of CARE collective on women's empowerment and how the gender composition of the collectives themselves and the leadership of the collectives moderates those effects. In order to do that, we retrieved a number of indicators from within each project's database. The specific indicators retrieved are as follows:

SDVC (Data collected from 2012 – current)
- Number of cows
- Breed of cows
- Litres of milk produced
- Quality of milk produced
- Income from milk
- Where milk sold
- Litres of milk consumed by HH members (detailed)
- Group membership list with gender
- Group leader list with gender

- Gender of person making spending decisions
- Gender of person making earning decisions
- Gender of person making internal household resource allocation decisions
- Gendered distribution of unpaid labor
- Experience with and attitudes towards domestic violence
- Sources of income for household
- Gendered distribution of access to inputs and extension services
- Gendered distribution of choices about inputs and extension services
- Women's mobility
- Women's comfort in public arena

Pathways (Data collected from 2013 – current)
- Number of recommended practices used
- Types of recommended practices used
- Amount and types of crops grown and harvested
- Productivity
- Profitability
- Group membership list with gender
- Group leader list with gender
- Gender of person making spending decisions
- Gender of person making earning decisions
- Gender of person making internal household resource allocation decisions
- Gendered distribution of help while planting, harvesting and weeding
- Experience with and attitudes towards domestic violence
- Sources of income for household
- Gendered distribution of access to inputs and extension services
- Gendered distribution of choices about inputs and extension services
- Women's mobility
- Women's comfort in public arena

*LinkUp (Data collected from 2015 – current)*
- Group membership list with gender
- Group leader list with gender
- Group savings
- Group return on investment
- Group return on assets
- Individual savings
- Individual accounts status
- Individual economic stability and satisfaction
- Gender of person making spending decisions
- Gender of person making earning decisions

- Gender of person making internal household resource allocation decisions
- Gendered distribution of household labour
- Sources of income for household
- Women's mobility
- Women's comfort in public arena

Pooled Approach

Advanced statistical techniques allow us to merge this data effectively and analyze it for global trends while still retaining the local nuances. When combining data from different sources, several technical points must be attended to in order to the resulting work to be accurate and reliable. The most vital of these is the comparability of the information obtained with different surveys.

We have taken a pooled approach to combining the databases. This is the ideal choice for this analysis since datasets are population independent – meaning that we have non-overlapping sampling frames. The pooled approach is illustrated below. In this approach, the samples from each of the surveys are combined into one large sample, with appropriate weight adjustments and an overall estimate is obtained from the pooled data
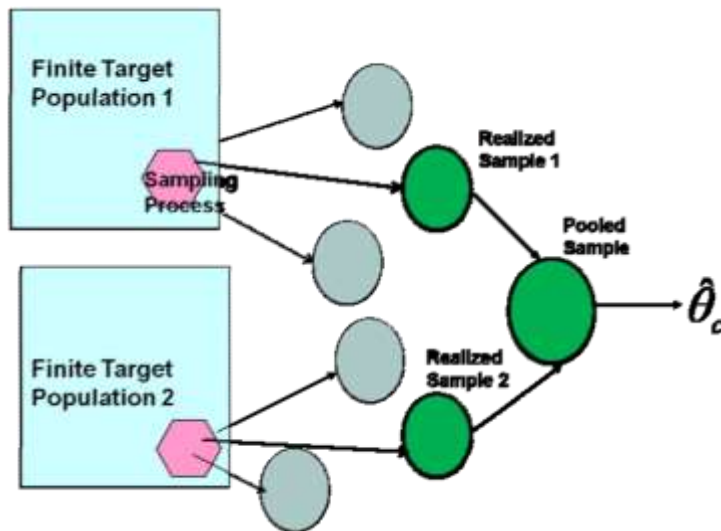
A simple random sample of a simple random sample is also a random sample. However, the surveys that these databases come from have a more complex design. Inclusion probabilities depend on y through many of the covariates. As a result of this, we cannot simply combine the estimates from each sample and hope that this represents an accurate estimate for the "superpopulation". (The superpopulation is the idea of the entire population that we are looking to estimate effects for – not simply the participants in one single CARE project.) Instead, we need to develop a system of weights which will give the greatest amount of weight to the most accurate estimates. We do this by extending our inference to the superpopulation using the joint processes of model and design.

Weights
The individual records from all available surveys have been combined and a set of new weights have been developed to merge the samples accurately. The pooled sample and the new weights have been used to build models that provide statistical estimates of trends and rates for the population as a whole.

For this project, we have chosen to use rescaling weights. If we used equal weights, this would be assuming that all the variation in estimates is due merely to sampling variation. We know that this is not the case for these datasets. Thus, pooled estimates need to be calculated taking into account as many sources of variation as possible. The weights we have used for rescaling the data was to rescale the weights by the factor of n_i/D_i where n is the sample size and D is the average design effect for the ith survey. This is motivated by the fact that this can yield the minimum variance estimates when the individual survey estimates are unbiased. This method is best when the population sizes are all large, which is the case for these three projects.

Standardization and normalization
When combining data from multiple data sources, it is necessary to preprocess the data to ensure the highest level of comparability, reliability and validity possible. The first step is to recode all the variables to ensure that they are all moving in the same direction – in other words, so that a higher number is a more positive event. Then, in order to standardize the data to allow for cross-project analysis, the categorical variables were derived to fit sets or normalized bins. The continuous variables were rescaled into z-scores. The z-score provides a simple measure by which different measures can be compared in terms of their deviation from the mean. This is often called standardization. Thus z is a measure of how far away a measurement is from the mean, measured in standard deviations. By recoding, standardizing and normalizing the data in this way, we have removed the dependency on units within each project. Measures of progress become comparable across projects and relative improvement on topics like the primary economic goal of each project become comparable – even though one has to do with milk production and one has to do with financial linkages, for example.

Hierarchical Mixed-Effects Models
A hierarchical Bayesian approach was developed to obtain model-based estimates derived from three types of direct project-level estimates. The differences between these three provide information about errors and nonresponse bias. Combining information based on a model that reflects both nonresponse and errors will improve the accuracy of the estimates. We assume that the population effect with a random error that stems only from the chance associated with the individual level of sampling error. Differences in the studies are in the power to detect the outcome of interest and each project is independent of the other projects. Here we assume asymptotically that the estimator of the treatment effect is a random variable with a normal distribution. We also assume that there is homogeneity of treatment effects across all the projects. We can then calculate the linear combinations of unbiased estimates that are themselves unbiased estimates and rely upon a weighted average from all the projects.

Then a series of models was developed to estimate the trends and effects of collectives across all three projects. The models were hierarchical mixed-effects models that incorporate the

dependent variable of interest predicted by involvement in collectives, composition of households, additional socio-economic variables as available, and group contextual variables. The models included appropriate numbers of dummy variables to account for multiple years and geographic locations. Thus the estimates derived from the model borrow strength across areas as well as time. In estimating variances, we used a Taylor linearization method that incorporates the survey weights, stratification, and clustering. The results of this series of models is displayed in this document.

Limitations
All research involving surveys have limitations. For this research, the two primary limitations are the reliability of the sampling frame and the comparability of specific questions. The samples of respondents who participated in the original surveys from the three projects are not representative of the entire population of the countries of the projects. This is because each CARE project is aimed at a specific subpopulation. As well, the sampling design used within each project is not perfectly random and therefore includes selection bias even within the target population. Therefore, even robust methods of combining these samples will not accurately represent the populations within each country. The generalizability of these results is limited by this factor.

The individual questions that are asked in each survey are not identical. For example, each survey does include a question about the gender of the household member who makes decisions about how the primary income is spent. However, each question is not worded in exactly the same way. Additionally, each question is asked in the local language, which contains nuances and social cues which aren't accounted for. This limits the perfect comparability of results across projects.

Finally, we used a Bayesian approach which requires analysts to input a suitable prior distribution. This method is best for our situation, since we do have previous data from each of the projects that we can use to develop estimates. However, as each project gathers more years of data, the overall estimates will become much more stable and have smaller standard errors.

Recommendations to CARE for future cross-project work
This analysis is only scratching the surface of what is possible by combining data from across projects. If slight changes were made to the way data was collected and stored, we could do much more in-depth research into how changes were affecting people with different characteristics as well as how things like the timing of projects, specific trainings and CARE infrastructure could be set up to achieve maximum impact most efficiently.

Some of our recommendations to set this up include:
1. Collecting data from the same people over time. It is much more effective to collect data from the same few people several times than it is to collect data from many different people once. Collecting data from the same people three times is really the minimum for a good cross-project database.

2. Creating a way to track people across time that does not include their name. If each project could create and use unique ID numbers for all project participants, it would save a tremendous amount of analytical time. This also helps with privacy concerns.

3. Use shorter surveys. By the time respondents reach the end of a some of the very long surveys used by CARE, your response error rate has dropped from 2% at the beginning of the survey to 12% at the end of the survey. We can help you use your current databases to pinpoint which specific indicators are providing you with the most important information specifically for your research questions. Then we can help you build your surveys around those.

4. Gather more directly aligned data. Create a list of questions on women's economic empowerment (or any other topic of common interest) that are used in very similar formats across project surveys. This is already very close to being accomplished organically – it would take only a small amount of fine tuning to get various projects collecting directly comparable data by asking the same questions.

5. Ask a few questions on respondent preferences. This would allow us to standardize data to local culture which would make it more comparable between cultures. For example, when asking about income levels, also ask about their satisfaction with current income. Or what their ideal income would be. This would allow us to adjust income variables both within cultures and across cultures. The same is true for information on decision making, access to inputs, education, and much more.

References
Béland Y. (2002), "Canadian Community Health Survey - Methodological Overview," *Health Reports* , Statistics Canada, Catalogue 82-003-X, Ottawa, 9-14.
Binder, D. A. and G. R. Roberts (2003), "Design-Based and Model-Based Methods for Estimating Model Parameters," in: R.L. Chambers and C.J Skinner eds., *Analysis of Survey Data*, Wiley, Chichester, 29-48.
Farrell, PJ, "Bayesian Inference for Small Area Proportions," Sankhya, Ser. B., 62, 402-416.
Gelfand, AE and Smith, AMF (1990) "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.
Korn, E. L. and B. I. Graubard (1999). *Analysis of Health Surveys*. Wiley, New York.
Lohr, SL. *Sampling: Design and analysis.* Pacific Grove (CA), 1999.
Schenker, N., Gentleman, J.F., Rose, D., Hing, E., and I.M. Shimizu (2002). "Combining estimates from complementary surveys: a case study using prevalence estimates from national health surveys of households and nursing homes," *Public Health Reports 2002*, 117, 393-407.
Schenker, N. and T.E. Raghunathan, (2007), "Combining Information from Multiple Surveys to Enhance Estimation of Measures of Health," *Statistics in Medicine, 26*, 1802-1811.
Thomas, S. (2007), "Combining Cycles of the Canadian Community Health Survey, *"Proceedings of Statistics Canada Symposium 2006: Methodological Issues in Measuring Population Health*, Statistics Canada, Catalogue 11-522-XIE, Ottawa.
Tjepkema, M. (2008), "Health Care Use Among Gay, Lesbian and Bisexual Canadians," *Health Reports, 19(1),* Statistics Canada, Catalogue 82-003, Ottawa, 53-64.