ICC Values in International Development: Evidence across Many Domains in sub-Saharan Africa

Sudhanshu Handa[1], Thomas de Hoop[2], Mitchell Morey[2] David Seidenfeld[2]

1. Department of Public Policy, UNC-Chapell Hill
2. American Institutes for Research

Abstract

There has been a sharp increase in the number of experimental studies evaluating development programs. However, recent evidence from the economics literature suggests reason for concern about the ability of current studies to detect meaningful effects with sufficient statistical power. The intraclass correlation is an important parameter for determining the statistical power of cluster-randomized trials. However, the parameter is rarely available to researchers planning a study until after the design is set and data are already collected. This paper takes an important step towards helping researchers to accurately estimate appropriate sample sizes for their clustered RCTs by presenting ICCs for a wide range of domains common for development research in Kenya, Malawi, Zambia, and Zimbabwe. Our results suggest that ICCs for common indicators in sub-Saharan Africa are lower than is commonly assumed in power calculations. However, we find higher ICC values for Kenya than in southern African countries across many domains, including educational enrollment and attendance, asset and livestock ownership, and fertility. At the same time, however, ICC values are universally low for indicators associated with nutrition and expenditures. For nutrition in particular the ICC values are lower than usually anticipated in power calculations, which suggests that major reductions in data collection costs can be achieved in impact evaluations that focus on nutrition.

Introduction

Policymakers and donors increasingly rely on evidence from randomized controlled trials (RCT) to make decisions about funding and scaling programs that aim to reduce poverty in low-and middle-income countries, for example in sub-Saharan Africa. The increased demand for RCTs and their use for allocating scarce resources and informing program design highlight the importance of properly designing each study. The high costs associated with implementing many RCTs only further supports the need to design them well. A well-designed study will have a sample size sufficiently large to detect meaningful effects, assessed by a power calculation with credible assumptions about the expected effect size, intraclass correlation, attrition rate and other key parameters. Unfortunately, however, many evaluations lack sufficient sample size to appropriately power their study. Ioannidis and Doucouliagos (2013) and Ioannidis, Stanley, & Doucouliagos (2017) argue that small sample sizes incapable of detecting the desired impacts cast doubt on much of the economic literature.

So what sample size is appropriate to detect meaningful effects for a randomized controlled trial? This usually depends on the unit of assignment. Often RCTs assign groups of units instead of individual units to the treatment or control condition, which can be defined as a cluster- randomized design. For these cluster-randomized designs determining the optimal sample size requires information about the intraclass correlation coefficient. The ICC is a statistic defined as the amount of correlation between units within and across clusters. However, ICCs are rarely known prior to collecting data for a study since they can vary by domain, geography, and size of the cluster, creating a serious challenge for designing an RCT with a sufficient sample size.

In fact very few studies show the value of ICCs and none cover domains for poor populations in developing countries. Hedges and Hedberg (2007) provides ICC values for reading and mathematics outcomes at each K-12 grade level in the United States. They present the ICCs for these important educational outcomes under various contexts to help provide a guide for future studies. Gulliford, Ukoumunne, and Chinn (1999) use a largescale community survey to provide a database of ICC values for various health outcomes of English adults. However, the ICCs listed in these papers are not relevant to determine the sample size of cluster-RCTs in sub-Saharan Africa leaving researchers guessing at what to use for an ICC when designing their study.

This paper takes an important step towards helping researchers to accurately estimate appropriate sample sizes for their clustered RCTs by presenting ICCs for a wide range of domains common for development research in sub-Saharan Africa. Specifically, we present ICCs for a wide range of commonly used development outcomes in eastern and southern Africa for labor constrained and very poor rural households, a common target population for programs aimed at reducing poverty in sub-Saharan Africa, such as cash transfer programs. We compare and contrast ICCs for several indicators within each domain, enabling researches to see how the ICC might vary by the measure they choose for a particular domain. We also compare and contrast ICCs across several countries to understand the stability of an ICC across different political and geographic contexts, yet within the same region of the world.

We are able to conduct this study because of the rare opportunity to access multiple datasets that share several important characteristics necessary for comparing ICCs including a similar target population of the sample, large sample size, similar time frame, documentation of clustered levels, common domains across studies and similar measures of indicators. We report ICCs for an array of common poverty and economic development indicators and domains including nutrition, health, education, food security, agriculture, household conditions, and economic wellbeing. The findings in this paper will enable researchers to design more accurate studies that produce higher quality evidence so that policymakers can make better decisions about how to allocate scarce resources to reduce poverty, ultimately benefitting the recipient of the aid program.

Intraclass Correlation Coefficients

The ICC helps determine the minimum detectable effect for a study, with larger ICCs making it more challenging to detect effects, all else equal. The exact ratio between standard errors for a cluster-level randomization and an individual-level randomization is

$$D = \sqrt{1 + (n-1)\rho}$$

Here n is the number of individuals per cluster and ρ is defined as the ICC (McKenzie (2012b)).

The ICC is then calculated using the formula

$$ICC = \frac{\rho_c}{\rho_c + \rho_i}$$

where ρ_c is the variation that occurs at the cluster level and ρ_i is the variation that occurs at the individual level. The value of the ICC indicates whether individuals are more similar across communities or within communities. The two limiting cases are where all variation occurs at the community level (ICC=1) or all variation occurs at the individual level (ICC=0). If all variation occurs at the community level ($\rho_c = 1$, $\rho_i = 0$), then all individuals within a community are essentially the same, as though each community had only one observation. If all of the variation occurs at the individual level ($\rho_c = 0$, $\rho_i = 1$), then each community is equivalent on average as if generated by a simple random sample at the population level. In this case, individual outcomes are unrelated to the community.

We calculate ICCs at the community level (group of villages) because it is the most common cluster-type for development economists. In a cluster-randomized trial, the best practice is to cluster standard errors at the unit of randomization (Abadie, Athey, Imbens, & Wooldridge, 2017). The programs were randomly assigned to either wards, communities, or villages because typically there are too few districts to conduct a sufficiently powered cluster-randomized trial with randomization at the district level. In our case the majority of the studies used either the ward or the community level as the unit of randomization because villages often vary in size, with some being too small to contain a sufficient number of households. In addition, randomization at the village level increases the risk of contamination.

Wards or communities of villages were the unit of randomization for the programs supporting this study's data and are common administrative units. Often, the next larger unit in sub-Saharan Africa is the district. Typically there are too few districts in a country to have sufficient numbers to randomize or the experiment would have to span the entire country.[1] The next smaller unit from community or Ward in most sub-Sahara African datasets is the village or neighborhood. It's possible to use villages as the unit of random assignment, however villages often vary in size, with some being too small to contain a sufficient number of households or it increases the likelihood of contamination between treatment and control groups since villages can flow into each other. For these reasons most cluster-randomized controlled trials in Eastern and Southern Africa randomly assign development programs at the ward or community level.

Description of Data

We use household-level data from five evaluations of large cash transfer programs in Kenya, Malawi, Zambia, and Zimbabwe. These studies evaluated large, national level cash transfer programs at the

[1] Muralidharan & Niehaus (2017) argue for the use of randomized controlled trials at scale, which may enable the use of randomization at the district level. However, at this almost all RCTs randomize at a lower level of aggregation.

household and individual level (REFERENCES). The datasets produced from these evaluations prove useful for this study of ICCs because several important characteristics are similar across the studies, including the target population of the sample, sample size, time frame, documentation of clustered levels, common domains and similar measures of indicators. The evaluated programs target similar populations of poor, labor constrained, food insecure, rural households that are often the beneficiaries of development programs. Thus, the ICCs are comparable across studies and the ICCs calculated from these data are likely to generalize to target populations of other development programs in the region. All studies include thousands of households with tens of thousands of individuals. The large sample enables us to calculate ICCs based on many households per cluster and many clusters per study. In addition, all studies were conducted around the same time, between 2007 and 2013, limiting the risk of any temporal effects that may reduce the comparability of ICCs across studies. The evaluation teams for each study recognized the nested nature of the data in their evaluation and accounted for it by documenting the link between each household and cluster, enabling us to calculate the ICC. Last, the studies investigated similar domains and used the same methods for measuring indicators for important economic development programs including food security, child nutrition, agricultural productivity, economic wellbeing, education, health, and housing conditions. Having similar measures on all of these domains enabled us to triangulate the estimated ICCs across countries, look for trends by domain, and learn what factors might contribute to higher ICCs by domain.

Table 1 summarizes the characteristics of each of the five evaluations whose data were used for this study. We use all households eligible for the program that were randomly selected as part of the treatment or control groups of each evaluation to calculate the ICCs. We use baseline data for the purpose of this study to ensure that ICCs are independent of the cash transfer program. However, we also calculate ICCs of several outcomes that were not recorded at baseline using endline data of the cash transfer program in Kenya. In the latter case, we calculated the ICCs based on households in the control group.

Table 1: Characteristics of the Data Sources

Program	**Country**	**Target Population**	**Year of Data Collection**	**Methodology**	**Name of clustering level**	**Number of Clusters**	**Number of Households**
Cash Transfer for Orphans and Vulnerable Children (CT-OVC)	Kenya	Poor households with orphans and vulnerable children	2007 (baseline) 2011 (endline)	Cluster-RCT	Sub-location	28	2,759
Social Cash Transfer Programme (SCTP)	Malawi	Ultra-poor, labor constrained	2013	Cluster-RCT	Village Cluster	29	3,531
Child Grant Cash Transfer Programme (CGP)	Zambia	Household with a child under five years old in three poor districts	2010	Cluster-RCT	CWAC	23	2,500

Multiple Category Targeting Grant (MCTG)	Zambia	Poor female- and elderly-headed households with disabled person	2011	Cluster-RCT	CWAC	39	3,078
Harmonized Social Cash Transfer (HSCT)	Zimbabwe	Food poor and labor constrained households	2011	Non-randomized	Ward	25	3,000

Several differences in the targeting strategy need to be considered in the interpretation of the results. The targeting strategy differed for each sample on the basis of the objectives of the cash transfer programs and the eligibility criteria of households. All households were, in part, selected on the basis of poverty levels: households from the Zambia CGP program come from three poor districts targeted by the program, whereas evaluation samples of the other four programs included households below context-specific poverty lines. In addition, programs included the following secondary criteria: the Kenya CT-OVC program targeted households with orphan and vulnerable children; the Malawi SCTP and Zimbabwe HSCT targeted labor constrained households; the Zambia CGP program targeted households with children under five; and the Zambia MCTG program included female- and elderly-headed households and households with disabled member. Within each household in Kenya, Malawi and Zambia and Zimbabwe, enumerators primarily collected data from the primary caregiver (typically female).

We present ICCs at the community levels that are most comparable across these four countries. These geographic levels are sub-locations (*mtaa mdogo*) in Kenya, village clusters in Malawi, community welfare assistance committees (CWACs) in Zambia, and wards in Zimbabwe.

Indicators:

This paper presents indicators from ten domains measured at either the individual or household level. Individual level indicators focus primarily on children and include nutrition, education, and material needs domains. Indicators at the household level focus on economic domains such as consumption, agricultural productivity, and livestock ownership, or poverty domains such as food security, poverty level, and household conditions. Below we describe how we construct indicators for analysis, and any differences in construction across datasets.

Household Level Indicators

Consumption indicators are calculated as total reported consumption within the household. We examine total consumption, consumption on health and hygiene (such as medicine, hospital fees, toiletries, cleaning expenses, etc.), education expenditures (tuition fees, transportation, uniforms, school meals, etc.), food consumption (all food and drink), food consumption excluding alcohol, and total consumption per capita. Consumption indicators vary slightly across countries due to context-specific items that are included for one country but not another, such as the consumption of regionally specific food.

The Zambian and Malawian samples include indicators that measure the total value of agricultural production, the value of food consumed from own production, and agricultural productivity, measured as the value of agricultural production divided by the amount of land used for farming.

We construct two indexes of household wealth using principle component factor analysis: one of household goods (whether the household owns a stove, television, radio, bed, etc.) and one of agricultural productive assets (including mower, plough, shovel, axe, etc.). We calculate ICCs of livestock ownership using the number of cattle, chicken, donkeys, goats, and pigs owned by the household.

Indicators of housing conditions include binary variables for whether the household has access to clean drinking water, plumbing, access to toilet facilities, a toilet on household premises, electricity, purchased roof materials, and purchased fuel for cooking and lighting.

We also include the Household Hunger Scale (HHS) and the Household Food Insecurity Access Scale (HFIAS) to measure food insecurity for Zambia and Zimbabwe samples. The HHS is measured using three questions on hunger experienced in the household, such as: “In the past four weeks, did you or any household member go to sleep hungry?” The scale is measured as the sum of responses “None”/”Sometimes”/”Often” coded as 0, 1, and 2. The HFIAS is the sum of responses to nine questions on food insecurity such as: “In the past month, have you worried that your household will not have enough food?”, and “In the past month, have you eaten a smaller meal than you felt you needed?” (Ballard et al (2011)). Responses of “Never”, “Rarely”, “Sometimes”, “Often” are coded from 0 to 3.

Individual level domains

Anthropometric measures are calculated for a randomly selected child within each household between 0-2 and 0-5 years of age. Using WHO guidelines, we calculate height-for-age, weight-for-age, and weight-for-height z scores and created indicators for children that are stunted, wasted, and underweight ($z < -2$), and those that are severely stunted, wasted, and underweight ($z < -3$).

We included indicators on school enrollment for a randomly selected child within each household between 7-14 and 15-18 years of age. Enrollment indicators include a binary variable for current enrollment, the number of school days the child attended class in the past week, and a binary variable for whether the child attended class every school day in the past week.

Indicators of female fertility are based on questions asked to all women in the household. If there were multiple women in a household, we prioritized the response of the household head’s spouse, or the household head, if it is a female-headed household. If the household did not include a female household head or wife of household head, we selected a woman age 18 to 30 at random; if none was available, we selected the response of a female household member age 31-49. We measured fertility indicators by including the age of first pregnancy, household size, number of biological children, and the number of biological children currently living in the household.

Results and Discussion

In this section we present the ICC values for indicators by domain and by country. We discuss the results in three ways:

1) Are the ICC values for an indicator consistent across countries?
2) Are the ICC values for a domain consistent across indicators?
3) What do the ICC values mean for the domain?

We examine the consistency of an indicator across countries to see if the ICC values are stable or if they vary across geography in the region. We then look at the consistency of ICC values for indicators within a domain to see if there is variation by the type of measure within the domain because that could help determine which indicators are the most relevant to use for power calculations. For example, we report ICCs for stunting, wasting, and underweight indicators within the child nutrition domain. If the ICCs for these different indicators vary from each other, than a study investigating child nutrition will have to make a decision about which ICC to use for the power calculation. However, if the ICCs for these nutrition indicators are all similar then the researcher can feel better about conducting power calculations based on one of the indicators. Last, we present the ICC values for each indictor so that researchers will have a better sense about how their study's power will be affected by a clustered design if one estimates the impact on that indicator, with higher ICCs providing lower power to detect effects, all else equal. We present the findings separately for each domain. The tables in this section present the ICC for each study/country by indicator with a 95% confidence interval for each ICC below it. We also present the weighted mean ICC for the indicator in the last column.

Individual Level Indicators

Child Nutrition

Table X.1 shows the ICC values for child nutrition indicators stunting, wasting, and underweight. We calculate ICCs for children under 5 years old and children under two years old. The ICCs for nutrition indicators are consistent for all comparisons. We only find differences of 0.03 between the min and max value for each of the nutrition indicators. Similarly, the ICC values for different nutrition indicators are consistent within the same country with stunting, wasting, and underweight all showing roughly the same ICC value. The ICC values for nutrition indicators are all very low with almost all of them under 0.05 and most under 0.03. Thus, there is very little difference in the variation of child nutrition within and between clusters in these countries. Clustering standard errors at the community level for these nutrition indicators will have a very small effect on statistical power as compared to individual random assignment, all else equal.

Table X.1 Child Nutrition ICC Values

Indicator	Kenya	Zimbabwe	Zambia (CGP)	Malawi	Average ICC
Stunted (0-24 months)	0.00	0.02 [0.00, 0.07]	0.02 [0.00, 0.04]	0.00 [0.00, 0.03]	0.01
Stunted (0-60 months)	0.02	0.02 [0.00, 0.04]	0.02 [0.00, 0.05]	0.02 [0.00, 0.04]	0.02
Severely stunted (0-24 months)	0.00	0.03 [0.00, 0.09]	0.01 [0.00, 0.03]	0.00 [0.00, 0.03]	0.01
Severely stunted (0-60 months)	0.02	0.01 [0.00, 0.04]	0.01 [0.00, 0.02]	0.00 [0.00, 0.01]	0.01
Height-for-age z-score (0-24 months)	0.00	0.03 [0.00, 0.09]	0.02 [0.00, 0.04]	0.01 [0.00, 0.04]	0.02
Height-for-age z-score (0-60 months)	0.04	0.02 [0.00, 0.05]	0.01 [0.00, 0.03]	0.01 [0.00, 0.03]	0.02
Wasted (0-24 months)	0.03	0.11*	0.01*	0.00	0.03

		[0.03, 0.18]	[0.00, 0.02]	[0.00, 0.03]	
Wasted (0-60 months)	0.01	0.06* [0.02, 0.10]	0.01 [0.00, 0.02]	0.00* [0.00, 0.01]	0.02
Severely wasted (0-24 months)	0.06	0.00 [0.00, 0.06]	0.00 [0.00, 0.01]	0.00 [0.00, 0.03]	0.01
Severely wasted (0-60 months)	0.00	0.04 [0.01, 0.07]	0.01 [0.00, 0.03]	0.00* [0.00, 0.01]	0.01
Weight-for-height z-score (0-24 months)	0.00	0.03 [0.00, 0.09]	0.01 [0.00, 0.03]	0.04 [0.00, 0.08]	0.02
Weight-for-height z-score (0-60 months)	0.01	0.03 [0.00, 0.07]	0.02 [0.00, 0.04]	0.02 [0.00, 0.04]	0.02
Underweight (0-24 months)	0.07	0.05 [0.00, 0.12]	0.00* [0.00, 0.01]	0.01 [0.00, 0.05]	0.02
Underweight (0-60 months)	0.03	0.01 [0.00, 0.04]	0.00 [0.00, 0.01]	0.00 [0.00, 0.01]	0.01
Severely underweight (0-24 months)	0.02	0.07 [0.00, 0.13]	0.00* [0.00, 0.01]	0.02 [0.00, 0.05]	0.02
Severely underweight (0-60 months)	0.01	0.06* [0.02, 0.09]	0.00* [0.00, 0.01]	0.01 [0.00, 0.02]	0.01
Weight-for-age z-score (0-24 months)	0.05	0.03 [0.00, 0.08]	0.00 [0.00, 0.02]	0.02 [0.00, 0.06]	0.02
Weight-for-age z-score (0-60 months)	0.05	0.02 [0.00, 0.05]	0.00* [0.00, 0.01]	0.00* [0.00, 0.01]	0.01

Child Education

Table X.2 shows the ICC values for the two most common education outcomes, enrolment and attendance. We break up each indicator by age group to represent primary and secondary school ages as supported by data from the cash transfer studies these data come from. We find consistent and very low ICC values for enrolment for both age groups across Zimbabwe, Zambia (both studies) and Malawi. Kenya however stands out with ICC values that are five to ten times higher. Kenya and Zimbabwe have similar ICC values for attendance that are much higher than Zambia. Unfortunately, we do not have ICC values for attendance in Malawi because the study did not include this indicator. Another pattern emerges where we find higher ICC values for primary school age children than secondary school age children. This pattern is especially pronounced in Kenya for both indicators and Zimbabwe for attendance. It unclear why Kenya has much higher ICC education values, especially for enrollment. The much higher ICC values for education outcomes in Kenya will have a large effect on the power to detect effects for these indicators, all else equal, especially for primary school age children for whom the ICC is estimated as 0.23.

Table X.2 ICCs for Child Education Indicators

Indicator	**Kenya**	**Zimbabwe**	**Zambia (CGP)**	**Zambia (MCTG)**	**Malawi**	**Average ICC**
	0.23*	0.02*	0.04	0.00*	0.03*	0.06

Currently enrolled (7-14 years)		[0.00, 0.04]	[0.01, 0.07]	[0.00, 0.01]	[0.01, 0.05]	
Currently enrolled (15-18 years)	0.13*	0.02 [0.00, 0.05]	0.06 [0.00, 0.12]	0.00* [0.00, 0.02]	0.02* [0.00, 0.04]	0.04
Days attended (7-14 years)	0.21	0.30* [0.22, 0.37]	0.02* [0.00, 0.05]	0.01* [0.00, 0.03]		0.16
Days attended (15-18 years)	0.12	0.16 [0.10, 0.22]	0.05 [0.00, 0.13]	0.04* [0.01, 0.09]		0.11
Full attendance in past week (7-14 years)	0.05	0.26* [0.19, 0.34]	0.04* [0.00, 0.07]	0.03* [0.00, 0.05]	0.03* [0.01, 0.05]	0.08
Full attendance in past week (15-18 years)	0.05	0.15* [0.10, 0.21]	0.13 [0.02, 0.23]	0.04* [0.00, 0.07]	0.05 [0.01, 0.09]	0.08

Fertility

Table X.3 shows the ICC values for fertility indicators, specifically age at first pregnancy and total number of children for all women in the study. These indicators have consistently similar ICC values both to each other within a country and across all five studies. Interestingly, the ICC values are quite low and stable with small confidence intervals. Thus, there is very little difference in the variation within a cluster as compared to across clusters for these indicators and the distribution of the population of the study closely resembles the distribution within any cluster. These results mean that the clustered aspect of a research design will not affect the power of the study to detect effects.

Table X.3 ICCs for Fertility

Indicator	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average ICC
Age of first pregnancy	0.00 [0.00, 0.04]	0.02 [0.00, 0.04]	0.01 [0.00, 0.03]	0.01 [0.00, 0.03]	0.01 [0.00, 0.02]	0.01
Number of Children	0.03 [0.00, 0.09]	0.02 [0.00, 0.04]	0.02 [0.00, 0.08]	0.01 [0.00, 0.03]	0.00* [0.00, 0.01]	0.01

Household Level Indicators

The following set of domains are measured at the household level

Poverty and Food Security

Table X.4 shows the ICC values for common poverty and food security indicators. Many of the cash transfer programs target eligible beneficiaries based on poverty level and food security, so the selection process for eligibility into the evaluated programs should generate a sample that is fairly consistent across clusters. Therefore we would expect to find fairly low ICCs for these indicators. Sure enough, as table X.4 shows, the ICC values for these indicators are fairly low and consistent across countries and across indicators within the same country.

Table X.4 ICC Values for Household Poverty and Food Security

Indicator	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average ICC

Below poverty line	0.06	0.02 [0.01, 0.03]	0.01 [0.00, 0.03]	0.01* [0.00, 0.01]	0.03 [0.01, 0.04]	0.02
Below food poverty line	0.05	0.04 [0.03, 0.06]	0.04 [0.01, 0.07]	0.02* [0.01, 0.04]	0.05 [0.02, 0.08]	0.04
Food Insecurity Access Scale		0.04* [0.02, 0.06]	0.16 [0.07, 0.24]	0.11 [0.06, 0.17]		0.09
Hunger Scale		0.05 [0.03, 0.07]	0.11 [0.04, 0.17]	0.06 [0.03, 0.10]		0.07
Meals per day		0.06 [0.04, 0.09]	0.09 [0.03, 0.15]	0.04 [0.02, 0.07]	0.08 [0.04, 0.12]	0.07

Livestock and Agricultural Assets

Most of the households in the 5 studies are rural, subsistence farming households. Table X.5 shows the ICC values for the number of livestock owned and an index of agricultural assets. With the exception of Kenya, we find fairly consistent ICC values across countries for each indicator making these fairly robust ICC values as a guideline for other studies. However we find more variation in ICC values across indicators. ICCs for the ownership of chickens and goats tends to be higher than for the ownership of pigs and cattle. Regardless, the ICC values for all of these indicators are below 0.1 and some below 0.05. The ICC for agricultural assets appears to have more variation across countries and also is a higher than for most livestock categories, thus potentially affecting statistical power more than for the livestock indicators.

Table X.5 ICC Values for Livestock and Agricultural Assets

Indicator	**Kenya**	**Zimbabwe**	**Zambia (CGP)**	**Zambia (MCTG)**	**Malawi**	**Average ICC**
Number of Pigs	0.01	0.03* [0.01, 0.04]	0.00* [0.00, 0.01]		0.01 [0.00, 0.01]	0.01
Number of Chickens	0.09	0.04 [0.02, 0.06]	0.07 [0.02, 0.11]	0.06 [0.03, 0.10]	0.01* [0.00, 0.02]	0.05
Number of Goats	0.22*	0.03* [0.01, 0.04]	0.07 [0.02, 0.11]	0.03* [0.01, 0.05]	0.02* [0.01, 0.04]	0.06
Number of Cattle	0.08*	0.04 [0.02, 0.05]	0.01* [0.00, 0.02]	0.00* [0.00, 0.01]	0.00* [0.00, 0.00]	0.02
Index of agricultural productive assets	0.10 [0.00, 0.21]	0.10 [0.06, 0.13]	0.02* [0.00, 0.04]	0.06 [0.03, 0.10]		0.07

Agricultural Production

Table X.6 shows the ICC values for several agricultural production indicators. We find a great range in ICC values both within indicators across countries and across indicators within a country. We do not have measures for all countries for these indicators so have fewer estimates to compare. Crop sales appears to be the most consistent across countries with the exception of Malawi which has a much higher ICC than in the other four studies.

Table X.6 Agricultural Production

Indicator	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average ICC
Crop sales	0.07 [0.00, 0.15]	0.00 [0.00, 0.19]	0.02 [0.00, 0.04]	0.05 [0.02, 0.08]	0.30* [0.19, 0.42]	0.04
Value of consumption from own production		0.06* [0.04, 0.08]	0.23* [0.11, 0.35]	0.14 [0.07, 0.20]	0.05* [0.02, 0.08]	0.11
Agricultural productivity ($/acre)			0.00* [0.00, 0.01]	0.11* [0.05, 0.16]	0.00* [0.00, 0.01]	0.04

Consumption

Table X.7 shows consumption and expenditure ICC values for different categories. These values are all relatively low with most under 0.1. We find fairly consistent values across studies for the same indicator; however, there is some variation across indicators. Expenditures on education and health have higher ICC values than for food consumption and overall consumption. These higher values could relate to the higher school enrolment values observed in the education section. Health might be similar to education in that some clusters have a clinic located near-by while others are further away, causing variation across clusters that does not exist as much within a cluster.

Table X.7 Consumption ICC Values

Indicator	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average ICC
Total consumption per capita	0.01*	0.03* [0.01, 0.04]	0.10 [0.04, 0.16]	0.04 [0.01, 0.06]	0.07 [0.03, 0.11]	0.05
Education expenditures per capita	0.10	0.05 [0.03, 0.07]	0.02* [0.00, 0.03]	0.02* [0.01, 0.04]	0.03 [0.01, 0.05]	0.04
Health and hygiene expenditures per capita	0.00*	0.03 [0.01, 0.04]	0.02 [0.00, 0.04]	0.10* [0.05, 0.15]	0.04 [0.01, 0.06]	0.04
Food consumption per capita	0.04	0.04 [0.02, 0.06]	0.11 [0.04, 0.18]	0.03* [0.01, 0.05]	0.07 [0.03, 0.10]	0.06
Food expenditures (excl. alcohol) per capita	0.04	0.04 [0.03, 0.06]	0.11 [0.04, 0.18]	0.03* [0.01, 0.05]	0.07 [0.03, 0.10]	0.06

Household Living Conditions

Table X.8 shows the ICC values for household living conditions, specifically household size, access to drinking water and access to a toilet or latrine. These indicators have little consistency across studies. These indicators also have some of the highest ICC values reported in this paper, meaning that they will dramatically affect power of a clustered randomized design. The ICC values for Malawi and the MCTG

Zambia study are fairly consistent within each of these studies and also on the moderate side as they hover around 0.1, however the ICC values within each of the other studies vary widely across the indicators.

Table X.8 ICC Values for Household Living Conditions

Indicator	**Kenya**	**Zimbabwe**	**Zambia (CGP)**	**Zambia (MCTG)**	**Malawi**	**Average ICC**
Household size	0.11 [0.00, 0.12]	0.05* [0.03, 0.07]	0.08 [0.03, 0.13]	0.09 [0.04, 0.14]	0.06 [0.03, 0.09]	0.07
Access to drinking water	0.39*	0.20 [0.15, 0.25]	0.21 [0.10, 0.33]	0.06* [0.03, 0.09]	0.10* [0.05, 0.16]	0.18
Access to toilet/latrine	0.38*	0.19 [0.14, 0.24]	0.46* [0.30, 0.63]	0.09* [0.04, 0.13]	0.11* [0.05, 0.17]	0.23
Owns toilet on HH premises	0.06*	0.20 [0.14, 0.25]	0.37* [0.22, 0.53]	0.21 [0.12, 0.30]	0.07* [0.03, 0.11]	0.18
Purchased roof	0.61*	0.20 [0.15, 0.25]	0.02* [0.00, 0.04]	0.04* [0.02, 0.07]		0.20
Purchased cooking fuel	0.27*	0.14 [0.10, 0.18]	0.09 [0.03, 0.14]	0.06* [0.03, 0.09]		0.13
Purchased lighting fuel	0.69*	0.16* [0.12, 0.21]	0.19* [0.09, 0.30]	0.23 [0.13, 0.32]		0.30
Has electricity			0.00 [0.00, 0.01]	0.00 [0.00, 0.01]		0.00

Conclusion

Recent evidence from the economics and other social science literature suggests reason for concern about the ability of current studies to detect meaningful effects with sufficient statistical power. Ioannidis, Stanley, & Doucouliagos (2017) show that most studies in economics are underpowered. In addition, Ioannidis (2005) shows that low power increases the likelihood of false positives. However, at the same time Banerjee et al. (2015) highlight the importance of conducting ex-ante power calculations in development economics research. Similarly. the number of impact evaluations in international development has increased significantly since 2009 (Brown et al. 2016). This promising trend has enabled development exconomists to increase the use of meta-analyses. As indicated by Ionnadis, Stanley, & Doucouliagos (2015) the use of meta-analyses enables international development researchers to estimate effect sizes of promising interventions while relying on several underpowered studies. Thus, despite the limited statistical power of current studies in economics, certain trends enable international development researchers to detect meaningful effects with sufficient precision.

This paper shows another reason to be more optimistic about the ability of development economics research to detect meaningful effects with sufficient precision. The findings about ICCs in sub-Saharan Africa suggest that ICCs in power calculations are often chosen too conservatively. Our results suggest that ICCs for common indicators in sub-Saharan Africa are lower than is commonly assumed in power calculations.

The ICC values we present are lower than 0.10 in almost all cases. For example, None of the ICCs for the indicators associated with consumption shows a value of higher than 0.10. Similarly, the ICCs for indicators associated with education and fertility are lower than 0.10 in each of the Southern African countries in our sample. Finally, ICC values of indicators associated with agriculture, disposal, food security, housing, and livestock are generally lower than 0.10. Moreover, many of the ICCs we present are lower than 0.05 or even 0.01.

The lower ICC values suggest that data collection costs could decrease significantly without losing the ability to detect small but meaningful effects of development programs with sufficient precision when we assume that all else remains equal. For example, an impact evaluation of a cash transfer program with random assignment at the community level would require a sample size of 800 treatment households and 800 control households in 40 treatment communities and 40 control communities to detect a treatment effect of 0.243 standardized mean differences on expenditures per capita when we would assume an ICC of 0.10. We would, however, only require a sample size of 540 treatment households and 540 control households across 54 communities to detect the same treatment effect when the ICC would be 0.05 as in the data we collected to determine the impact of the Zambian social cash transfer program. The required sample size for detecting this impact would decrease further to 160 treatment and 160 control households when the ICC would be 0.005, which is the ICC value for levels of stunting between 2 and 5 years old in the evaluation of the Zambian social cash transfer program.

Nonetheless, we find important differences in ICC values across countries. Specifically, we find higher ICC values for Kenya than in southern African countries across many domains, including educational enrollment and attendance, asset and livestock ownership, and fertility. These differences show the importance of accounting for differences in contextual characteristics when conducting power calculations. It is not sufficient to rely on ICCs from different contexts for conducting power calculations. Instead, it will be important to widely document ICCs from different contexts across many indicators that are commonly used for impact evaluations in international development.

At the same time, however, ICC values are universally low for indicators associated with nutrition and expenditures. For nutrition in particular the ICC values are lower than usually anticipated in power calculations, which suggests that major reductions in data collection costs can be achieved in impact evaluations that focus on nutrition. Although the ICCs for expenditures are not as low as for nutrition, our results also suggest that it may be feasible to reduce sample sizes for estimating the impact of international development programs that focus on expenditures. Arguably, however, reducing sample sizes will only be appropriate when effect sizes used in power calculations are selected appropriately.

However, caution is needed in the interpretation of our results. Although ICCs may be lower than is commonly anticipated in power calculations, there is less reason for optimism about other parameters. For example, McKenzie & Woodruff (2014) show that impact evaluations of business training and entrepreneurship evaluations are systematically underpowered because researchers make too optimistic assumptions about take-up rates. Similarly, anticipated effect sizes are routinely overestimated in power calculations (Reference).

These findings suggest that future research should focus on estimating other parameters that are relevant for power calculations in a wide range of settings. For example, researchers could focus on summarizing the effect sizes of common interventions in international development, for example through the use of meta-analyses. In addition, future research could focus on summarizing take-up and attrition rates across interventions and contexts.

Finally, we need to remain careful in the interpretation of the results because the ICCs we summarize are specific to certain contexts in sub-Saharan Africa with labor-constrained households. Although this population can be considered relevant for many development interventions, it will be important to summarize ICCs from large sample across a wider range of contexts. Such evidence would enable researchers to conduct power calculations with more accurate assumptions, which could in turn result in higher-quality impact evaluations in international development.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *The Journal of economic perspectives* 24.2 (2010): 3-30.

Ballard, Terri; Coates, Jennifer; Swindale, Anne; and Deitchler, Megan. Household Hunger Scale: Indicator Definition and Measurement Guide. Washington, DC: Food and Nutrition Technical Assistance II Project, FHI 360.

Banerjee, Abhijit; Karlan, Dean; and Johnathan Zinman (2015) "Six randomized evaluations of microcredit: introduction and further steps," American Economic Journal: Applied Economics, Vol. 7.1, pp. 1-21.

Button, Katherine S, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munaf (2013) :Power failure: why small sample size undermines the reliability of neuroscience," Nature Reviews Neuroscience, Vol. 14, pp. 365{376.

Cameron, A. Colin, and Douglas L. Miller. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources* 50.2 (2015): 317-372.

Campbell, Marion K., Jill Mollison, and Jeremy M. Grimshaw. "Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size." *Statistics in medicine* 20.3 (2001): 391-399.

Cohen, Jacob (1977) Statistical power analysis for the behavioral sciences (rev: Lawrence Erlbaum Associates, Inc.

Duflo, Esther, Rachel Glennerster, and Michael Kremer (2007) \Using randomization in development economics research: A toolkit," Handbook of development economics, Vol. 4, pp. 3895{3962.

Gulliford, Martin C., Obioha C. Ukoumunne, and Susan Chinn. "Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994." *American Journal of Epidemiology* 149.9 (1999): 876-883.

Hedges, Larry V and Eric C Hedberg (2007) \Intraclass correlation values for planning group-randomized trials in education," Educational Evaluation and Policy Analysis, Vol. 29, pp. 60{87.

Ioannidis, John PA, Tom D. Stanley, and Hristos Doucouliagos. "The power of bias in economics research." The Economic Journal (2017).

Ioannidis, John, and Chris Doucouliagos. "What's to know about the credibility of empirical economics?." *Journal of Economic Surveys* 27.5 (2013): 997-1004.

Ioannidis, John PA. "Why most published research findings are false." *PLos med* 2.8 (2005): e124.

McKenzie, David. "Tools of the Trade: Intra-cluster Correlations." Impact Evaluations. World Bank, 02 Dec. 2012.

McKenzie, David; and Christopher Woodruff (2014) What are we learning from business training and entrepreneurship evaluations around the developing world? Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/22564/wbro_29_1_48.pdf;sequence=1

Sudhanshu Handa. Impacts of Social Cash Transfers in sub-Saharan Africa: Evidence from the Transfer Project. Presented at *Social Protection: Building Effective and Sustainable Systems for Equitable Growth*, Arusha, Tanzania, December 15-17, 2014

Wooldridge, Jeffrey M (2010) Econometric analysis of cross section and panel data: MIT press.

Tables

Table 2: Expenditures

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Total expenditures per capita	0.013	0.014	0.049	0.035	0.068	0.036
Education expenditures	0.098	0.013	0.009	0.019	0.029	0.034
Health and hygiene expenditures	0.000	0.009	0.009	0.094	0.035	0.029
Food expenditures in total	0.042	0.015	0.052	0.030	0.065	0.041
Food expenditures (excluding alcohol)	0.042	0.016	0.052		0.065	0.044

Table 3: Agriculture

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Value of agricultural production			0.000	0.048	0.302	0.117
Crop sales	0.067[e]					0.067
Value of consumption from own production			0.203	0.132	0.053	0.129
Agricultural productivity ($/acre)			0.000	0.105	0.003	0.036
Index of agricultural productive assets	0.101[e]		0.001	0.062		0.055
Index of household assets	0.761		0.019	0.040		0.273

Table 4: Food security

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Household Food Insecurity Access Scale		0.012	0.120	0.106		0.079
Household Hunger Scale		0.016	0.075	0.058		0.050

Average meals per day	0.010	0.050	0.041	0.078	0.045

Table 5: Stunting

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Stunted (0-24 months)	0.000	0.020	0.000		0.000	0.007
Stunted (0-60 months)	0.021	0.012	0.005		0.020	0.014
Severely stunted (0-24 months)	0.000	0.000	0.000		0.000	0.000
Severely stunted (0-60 months)	0.020	0.002	0.001		0.002	0.006
Height-for-age z-score (0-24 months)	0.000	0.027	0.000		0.008	0.009
Height-for-age z-score (0-60 months)	0.038	0.015	0.004		0.015	0.018

Table 6: Underweight

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Underweight (0-24 months)	0.069	0.000	0.000		0.011	0.020
Underweight (0-60 months)	0.033	0.000	0.000		0.003	0.009
Severely underweight (0-24 months)	0.021	0.000	0.000		0.016	0.009
Severely underweight (0-60 months)	0.005	0.016	0.000		0.010	0.008
Weight-for-age z-score (0-24 months)	0.054	0.000	0.000		0.021	0.019
Weight-for-age z-score (0-60 months)	0.053	0.000	0.000		0.000	0.013

Table 7: Wasting

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Wasting (0-24 months)	0.025	0.000	0.003		0.000	0.007
Wasting (0-60 months)	0.006	0.001	0.004		0.000	0.003
Severely wasting (0-24 months)	0.061	0.000	0.000		0.002	0.016
Severely wasting (0-60 months)	0.000	0.009	0.003		0.000	0.003
Weight-for-height z-score (0-24 months)	0.000	0.000	0.003		0.035	0.009
Weight-for-height z-score (0-60 months)	0.010	0.005	0.006		0.024	0.011

Table 8: Enrollment

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi	Average
Currently enrolled (7-14 years)	0.233	0.008	0.018	0.000	0.0326	0.058
Currently enrolled (15-18 years)	0.129	0.004	0.013	0.000	0.0199	0.033
Days attended (7-14 years)	0.213	0.027	0.008	0.011		0.065
Days attended (15-18 years)	0.124	0.018	0.000	0.046		0.047
Full attendance in past week (7-14 years)	0.047	0.042	0.013	0.026	0.0333	0.032
Full attendance in past week (15-18 years)	0.051	0.024	0.028	0.039	0.0519	0.039

Table 9: Livestock

ICC (Ward level)	Kenya	Zimbabwe	Zambia (CGP)	Zambia (MCTG)	Malawi
Household size	0.108	0.010	0.039	0.084	0.059
Age of first pregnancy	0.000[e]	0.003	0.011	0.012	0.005
Biological children living in household	0.000[e]	0.005	0.014	0.038	0.000
Biological children in total	0.029[e]	0.006	0.028	0.009	0.000