

Intraclass Correlations Values in International Development: Evidence Across Commonly Studied Domains in sub-Saharan Africa

Evaluation Review
2023, Vol. 47(5) 786–819
© The Author(s) 2023





Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0193841X231154714

journals.sagepub.com/home/erx



David Seidenfeld¹, Sudhanshu Handa²,
Thomas de Hoop³ , and Mitchell Morey 

Abstract

The sharp increase in the number of experimental studies evaluating development programs raises the need for accurate intraclass correlations (ICC) to conduct power calculations so that researchers can design 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. 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 working in sub-Saharan Africa to accurately estimate appropriate sample sizes for their clustered RCTs. The study draws from rich data sets in Kenya, Malawi, Zambia, and Zimbabwe. We present ICCs for a wide range of domains common for development research. Our results suggest that ICCs for commonly studied indicators in sub-Saharan Africa are lower than is often

¹International Development Division, American Institutes for Research, Arlington, VA, USA

²Department of Public Policy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and Institute Fellow, AIR, Nairobi, Kenya.

³Principal Economist, American Institutes for Research, Washington, DC, USA

Corresponding Author:

Mitchell Morey, Senior Economist American Institutes for Research 1400 Crystal Drive 10th Floor Arlington, VA, USA.

Email: mmorey@air.org

assumed in power calculations. ICC values are especially low for indicators associated with child nutrition and food security, suggesting that cluster-RCTs might be a viable design even when faced with limited budgets because sample size requirements are not much different from an individual random assignment design.

Keywords

intracluster correlation, Study design, sub-Saharan Africa, Randomized controlled trial

Introduction

Policymakers and donors increasingly rely on evidence from randomized controlled trials (RCTs) 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 support 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, many evaluations lack a sufficient sample size to appropriately power their study, suggesting the need for power calculations with more credible assumptions. [Button et al \(2013\)](#), [Ioannidis and Doucouliagos \(2013\)](#), [Isaakidis and Ioannidis \(2003\)](#), and [Ioannidis et al. \(2017\)](#) argue that small sample sizes incapable of detecting desired impacts cast doubt on much of the economics literature. One way to account for this challenge is by having more information on intracluster correlations before the start of the study. However, researchers must often design a study without the benefit of existing data on the study sample, limiting their ability to determine intracluster correlations based on existing data. For example, a study may include randomization without baseline data. Alternatively, program implementation may start immediately upon completion of baseline data collection, preventing researchers to use intracluster correlations from the baseline for power calculations.

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 is defined as a cluster-randomized design. Studies with cluster-randomized designs should account for the clustering by using cluster-robust standard errors that account for a lack of independence across

observations (Duflo et al., 2007, Cohen, 1977, Abadie et al., 2022, Angrist & Jörn-Steffenm, 2010, Wooldridge & Jeffrey, 2010). For these cluster-randomized designs, determining the optimal sample size requires information about the intraclass correlation coefficient (ICC). The ICC is a statistic defined as the amount of correlation between units within and across clusters. Unfortunately, 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 challenge for designing a cluster RCT with a sufficient sample size.

There are a limited number of studies that provide the value of ICCs in developing countries and none that provide a collection of ICCs across domains. Three studies in particular provide two-level ICC values for children's learning outcomes when children are clustered in schools (Spybrook & Kelcey, 2016; Kelcey et al., 2016, Kelcey et al., 2017). These studies find that learning outcomes are highly clustered within schools but that the degree of clustering varies widely across countries. They provide tools to researchers to address the existing evidence that schools are heterogeneous and interventions have different effects in each school (Heneveld & Craig, 1996). Geyer et al. (2016) provide a similar set of design parameters, but their information is limited to agricultural outcomes in Mozambique. Zopluoglu (2012) estimates and compares ICCs across different countries and finds substantial differences, but this study focuses primarily on OECD countries.

Most research on ICCs focuses on the United States. Hedges and Hedberg (2007) provide ICC values for reading and mathematics outcomes at each K-12 grade level in the United States. They also show heterogeneity by state in a follow-up study (Hedges & Hedberg, 2013). They present the ICCs for educational outcomes under various contexts to help provide a guide for future studies. Similarly, Jacob et al. (2010) provide ICC values from education outcomes as well as a limited number of health outcomes. Westine et al. (2013) further develop the evidence for ICCs in the U.S. context, reporting ICCs for different subjects and different grade levels within the state of Texas. Gulliford et al. (1999) use a large-scale community survey to provide a database of ICC values for various health outcomes of English adults. However, the ICCs from studies in the United States are less relevant to determine the sample size of cluster-RCTs in sub-Saharan Africa leaving researchers guessing at what value 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 frequently 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 researchers to see how the ICC might vary by the measure they choose for a particular domain. We also compare and contrast ICCs for several countries to understand the stability of an ICC across different political and geographic contexts, yet within the same region of the world. Finally, we present design effects and examine how much ICCs change after adjusting for covariates.

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 recipients of development programs.

Intraclass Correlation Coefficients

The ICC helps to determine the minimum detectable effect for a study, with larger ICCs making it more challenging to detect effects, all else equal. The design effect to adjust the sample size in a cluster-level randomization relative to an individual-level randomization is characterized as in equation (1)

$$D = 1 + (n - 1)\rho \quad (1)$$

Here, D is the ratio of the variance of a cluster-randomized controlled trial and an individual-level randomized controlled trial. It is a function of n , which is the average number of individuals per cluster and ρ , which is defined as the ICC (Ukoumunne, 2002). An important assumption underlying this equation is that cluster sizes are consistent. Varying cluster sizes would further increase the design effect. Furthermore, the equation shows that large clusters can lead to large design effects even when the ICC is small. The ICC is then calculated using the formula

$$\rho = \frac{v_c}{v_c + v_i}$$

Here, v_c is the variation that occurs at the cluster-level and v_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 ($v_c = 1, v_i = 0$), 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 ($v_c = 0, v_i = 1$), then a cluster-randomized design approximates a simple random sample at the population level. In this case, individual outcomes are uncorrelated within the community.

In a cluster-randomized trial, the best practice is to cluster standard errors at the unit of randomization (Abadie et al., 2022). Wards or communities of villages were the unit of randomization for the cash transfer programs evaluated using this study's data and are common administrative units in sub-Saharan Africa. Often, the next larger unit in sub-Saharan Africa is the district. Typically, there are too few districts in a country to achieve sufficient statistical power with district-level random assignment, unless experiments span the entire population of a large country. The next smaller unit in most sub-Saharan African datasets after community or Ward is typically the village. It's possible to use villages as the unit of random assignment, but they often vary in size, with some being too small to contain a sufficient number of households. Choosing villages as the unit of random assignments may also increase 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 programs at the ward or community level.

This paper presents ICCs with and without covariate adjustment. Following Hedges & Hedberg (2007), we use the mixed command in Stata for linear model analysis to calculate ICCs with and without covariate adjustment.

We also calculated covariate-adjusted ICC estimates because adding covariates can increase statistical power substantially. Controlling for covariates could thus justify using lower ICCs in study planning [if the difference with ICCs without adjustment is substantial (Shoukri et al., 2013, Murray & Blitstein, 2003; McKenzie, 2012)]. To calculate covariate-adjusted ICC estimates, we selected a standard set of covariates regularly applied in cash transfer studies: household size and the age, gender, and primary school completion status of the survey respondent. While other covariates might also be relevant for certain outcomes or in certain settings, these four covariates apply to all outcomes in this study.

Description of Data

We use household and individual-level data from five evaluations of cash transfer programs in Kenya, Malawi, Zambia, and Zimbabwe, which we summarize in Table 1. These studies evaluated large, national level cash

transfer programs by estimating impacts at the household and individual level. The datasets produced from these evaluations prove useful for this study 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 survey tools applied in the five studies generally drew from the same survey modules, which further enhances the comparability of ICCs derived from these datasets.¹ The tools themselves include validated modules such as UNICEF's Multiple Indicator Cluster Survey as well as validated modules for consumption and asset ownership. The five programs target similar populations of poor, labor constrained, food insecure, rural households that are often the main beneficiaries of cash transfer and other development programs. Thus, the ICCs are comparable across studies and the ICCs calculated from these data may generalize to target populations of other development programs in the region. All five of the 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. 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, examine 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.

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 program 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

Table 1. Characteristics of the cash transfer programs and data sources.

Program	Country	Target population	Year of data collection	Methodology	Name of clustering Level	Average cluster size	# of clusters	Total # of sampled households
Cash Transfer for Orphans and Vulnerable Children(CT-OVC)	Kenya	Poor households with orphans and vulnerable children	2007	Cluster-RCT	Sub-location	5855 people; 33.9 sq miles	28	2759
Social Cash Transfer Programme (SCTP)	Malawi	Ultra-poor, labor constrained	2013	Cluster-RCT	Village cluster	7090 people; 20.5 sq miles	29	3531
Child Grant Cash Transfer Programme(CGP)	Zambia	Household with a child under five years old in three poor districts	2010	Cluster-RCT	CWAC	1811 people; 81.1 sq mi	44	2500
Multiple Category Targeting Grant (MCTG)	Zambia	Poor female- and elderly-headed households with disabled person	2011	Cluster-RCT	CWAC	1811 people; 81.1 sq mi	42	3078
Harmonized Social Cash Transfer (HSCT)	Zimbabwe	Food poor and labor constrained households	2011	Non-randomized	Ward	10,741 people; 125.7 sq miles	91	3000

under five; and the Zambia MCTG program included female- and elderly-headed households and households with disabled members. 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. Each of the studies used a similar sampling approach in which the total sample was divided evenly across clusters.

Indicators

This paper presents indicators from ten domains measured at either the household or individual level. Indicators at the household level focus on economic domains such as consumption, agricultural productivity, and livestock ownership, or poverty domains including food security, poverty level, and household conditions. Individual-level indicators focus primarily on children and include nutrition, education, and material needs domains. Next, 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, which is measured as the value of agricultural production divided by the amount of land used for farming.

We construct two indexes of household wealth using principal 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

Child nutrition measures are calculated for a randomly selected child within each household between 0–2 and 0–5 years of age. Using WHO guidelines, we calculated 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 aged 18–30 years old at random; if none was available, we selected the response of a female household member aged 31–49 years old. We measured fertility indicators by including the age of first pregnancy, 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. In each case, we calculate the ICCs as described above. Our findings

provide the key aspects of reporting ICCs (Campbell et al., 2004). 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 geographies 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, then a study investigating child nutrition will have to decide about which ICC to use for the power calculation. However, if the ICCs for these nutrition indicators are all similar, then the researcher can choose to conduct power calculations based on one of the indicators. Last, we present the ICC values for each indicator so that researchers will understand better 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. These results represent a conservative upper-bound on ICC values that an applied researcher can use in designing their study. For a research design that will include other explanatory variables, covariate adjustment of ICCs can justify using even lower ICCs in study planning (Murray & Blitstein, 2003; McKenzie, 2012) if the difference between adjusted and unadjusted ICCs is meaningful.

Tables 2, 3, 4, 5, 6, 7, 8, and 9 present unadjusted ICC estimates alongside the covariate-adjusted ICCs. We consider the unadjusted values as our primary set of findings and the ones most relevant for most applied researchers. These unconditional results come from the simplest analytical design and reflect the information most readily available to researchers, who often plan study sample sizes prior to collecting any data. Furthermore, we find that the covariate-adjusted model largely provides similar results, with the ICCs only reducing by about 0.01 to 0.03 in general in comparison with the ICCs that do not apply covariate adjustment.

Throughout our discussion, we provide estimates of typical design effects for that domain. In doing so, we assume an average cluster size of 63 households, which is typical for the studies producing the data in this paper.

Design effects produced using this assumption are larger than those with a research design using a smaller number of households per cluster.

To further aid in interpreting the results, we present descriptive statistics for all outcomes at the country-level in Annex A. These results provide the mean value and the standard deviation. These values are important because an outcome with very little variation may have more precisely estimated ICCs.

Individual-Level Indicators

Child Nutrition. Table 2 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 across indicators and across countries. We find a maximum difference of 0.03 between the minimum and maximum value for each of the ICCs 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 low with almost all of them under 0.05 and most under 0.03. Thus, there is 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 only have a small effect on statistical power compared to individual random assignment, all else equal, making a clustered RCT an attractive design for nutrition studies compared to outcomes for which ICC values are higher. Estimates of design effects suggest that a typical design effect is 2.24 for studies with similar cluster sizes to those in the evaluations we report on.

Child Education. Table 3 shows the ICC values for the two most common education outcomes, enrollment, 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 low ICC values for enrollment for both age groups across Zimbabwe, Zambia (both studies) and Malawi. One possible explanation for the relatively low ICC values is that the programs all target vulnerable populations. The ICCs may thus be lower than ICCs estimated from a national census of child outcomes. Kenya does, however, stand out, especially for enrollment, with ICC values that are five to ten times higher. Part of the difference that leads to higher ICCs for Kenya may be the more urban nature of that dataset, where enrollment rates are much higher in urban settings than rural ones. Kenya and Zimbabwe have similar ICC values for attendance. These values are much higher than the ICC values for attendance in Zambia. Unfortunately, we do not have ICC values for

Table 2. Child nutrition ICC values.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Malawi		Average ICC	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Stunted (0–24 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01 [0.00,0.07]	0.01 [0.00,0.08]	0.02 [0.00,0.05]	0.03 [0.00,0.06]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00	0.01
Stunted (0–60 months)	0.02 [0.00,0.05]	0.05 [0.00,0.12]	0.02 [0.00,0.04]	0.01 [0.00,0.04]	0.03 [0.01,0.05]	0.03 [0.00,0.05]	0.02 [0.00,0.04]	0.02 [0.00,0.04]	0.02	0.03
Severely stunted (0–24 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.02 [0.00,0.08]	0.06 [0.00,0.15]	0.01 [0.00,0.03]	0.02 [0.00,0.05]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01	0.02
Severely stunted (0–60 months)	0.02 [0.00,0.05]	0.02 [0.00,0.05]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.01 [0.00,0.02]	0.01 [0.00,0.02]	0.00 [0.00,0.01]	0.00 [0.00,0.01]	0.01	0.01
Height-for-age z-score (0–24 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.02 [0.00,0.08]	0.02 [0.00,0.08]	0.01 [0.00,0.04]	0.02 [0.00,0.05]	0.01 [0.00,0.04]	0.01 [0.00,0.04]	0.01	0.01
Height-for-age z-score (0–60 months)	0.03 [0.00,0.08]	0.04 [0.00,0.10]	0.02 [0.00,0.05]	0.01 [0.00,0.04]	0.02 [0.00,0.04]	0.02 [0.00,0.04]	0.01 [0.00,0.03]	0.02 [0.00,0.04]	0.02	0.02
Wasted (0–24 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.03 [0.00,0.10]	0.08 [0.00,0.18]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01	0.02
Wasted (0–60 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.06 [0.02,0.10]	0.01 [0.00,0.06]	0.01 [0.00,0.02]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.00 [0.00,0.01]	0.02	0.00

(continued)

Table 2. (continued)

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Malawi		Average ICC	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Severely wasted (0–24 months)	0.05 [0.00,0.14]	0.04 [0.00,0.14]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.03	0.01
Severely wasted (0–60 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.03 [0.00,0.06]	0.00 [0.00,0.00]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01	0.00
Weight-for-height z-score (0–24 months)	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01 [0.00,0.07]	0.01 [0.00,0.08]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.03 [0.00,0.10]	0.00	0.01
Weight-for-height z-score (0–60 months)	0.00 [0.00,0.03]	0.00 [0.00,0.00]	0.03 [0.00,0.06]	0.00 [0.00,0.03]	0.02 [0.00,0.04]	0.01 [0.00,0.03]	0.02 [0.00,0.05]	0.02 [0.00,0.05]	0.01	0.01
Underweight (0–24 months)	0.05 [0.00,0.14]	0.00 [0.00,0.00]	0.03 [0.00,0.09]	0.06 [0.00,0.15]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01 [0.00,0.05]	0.01 [0.00,0.05]	0.02	0.02
Underweight (0–60 months)	0.02 [0.00,0.05]	0.01 [0.00,0.04]	0.01 [0.00,0.04]	0.01 [0.00,0.04]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.00 [0.00,0.01]	0.00 [0.00,0.01]	0.01	0.01

(continued)

Table 2. (continued)

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Malawi		Average ICC	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Severely underweight (0-24 months)	0.01 [0.00,0.07]	0.04 [0.00,0.13]	0.05 [0.00,0.12]	0.13 [0.03,0.23]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01 [0.00,0.05]	0.01 [0.00,0.06]	0.02	0.05
Severely underweight (0-60 months)	0.01 [0.00,0.03]	0.02 [0.00,0.06]	0.06 [0.02,0.10]	0.06 [0.01,0.11]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.01 [0.00,0.03]	0.01 [0.00,0.02]	0.02	0.02
Weight-for-age z-score (0-24 months)	0.05 [0.00,0.13]	0.01 [-0.10,0.12]	0.03 [0.00,0.09]	0.08 [0.00,0.16]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.01 [0.00,0.06]	0.00 [0.00,0.00]	0.02	0.02
Weight-for-age z-score (0-60 months)	0.05 [0.00,0.09]	0.03 [0.00,0.08]	0.03 [0.00,0.05]	0.02 [0.00,0.06]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.01 [0.00,0.06]	0.02	0.02

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Table 3. ICCs for child education indicators.

Indicator	Kenya			Zimbabwe			Zambia (CGP)			Zambia (MCTG)			Malawi			Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	
Currently enrolled (7–14 years)	0.34 [0.21,0.47]	0.34 [0.14,0.55]	0.02 [0.00,0.03]	0.02 [0.02,0.02]	0.04 [0.01,0.07]	0.03 [0.00,0.06]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.03 [0.01, 0.05]	0.03 [0.01, 0.05]	0.08 [0.01, 0.06]	0.08 [0.00,0.00]	0.03 [0.01, 0.05]	0.03 [0.01, 0.05]	0.08 [0.01, 0.06]	0.08 [0.00,0.00]	
Days attended (7–14 years)	0.31 [0.18,0.43]	0.23 [0.04,0.42]	0.46 [0.37,0.55]	0.43 [0.43,0.44]	0.02 [0.00,0.05]	0.03 [0.00,0.07]	0.01 [0.00,0.03]	0.00 [0.00,0.02]	0.03 [0.00,0.04]	0.03 [0.00,0.07]	0.03 [0.00,0.07]	0.00 [0.00,0.02]	0.00 [0.00,0.03]	0.00 [0.00,0.02]	0.20 [0.01, 0.06]	0.14 [0.00,0.02]	
Days attended (15–18 years)	0.15 [0.06,0.24]	0.08 [0.01,0.15]	0.18 [0.12,0.24]	0.19 [0.18,0.19]	0.04 [0.00,0.11]	0.04 [0.00,0.11]	0.04 [0.01,0.08]	0.05 [0.00,0.09]	0.04 [0.01, 0.06]	0.04 [0.01, 0.06]	0.04 [0.01, 0.06]	0.05 [0.00,0.09]	0.04 [0.01, 0.06]	0.04 [0.01, 0.06]	0.10 [0.01, 0.06]	0.13 [0.00,0.04]	
Full attendance in past week (7–14 years)	0.05 [0.02,0.09]	0.02 [0.00,0.05]	0.31 [0.24,0.39]	0.30 [0.30,0.31]	0.04 [0.00,0.07]	0.05 [0.00,0.09]	0.02 [0.00,0.05]	0.02 [0.00,0.04]	0.04 [0.01, 0.06]	0.04 [0.01, 0.06]	0.04 [0.01, 0.06]	0.02 [0.00,0.04]	0.02 [0.01, 0.06]	0.02 [0.01, 0.06]	0.09 [0.01, 0.06]	0.10 [0.00,0.04]	
Full attendance in past week (15–18 years)	0.05 [0.01,0.10]	0.04 [0.00,0.08]	0.17 [0.11,0.23]	0.19 [0.18,0.19]	0.13 [0.03,0.24]	0.14 [0.02,0.26]	0.04 [0.00,0.07]	0.03 [0.00,0.07]	0.02 [0.01, 0.04]	0.02 [0.01, 0.04]	0.02 [0.01, 0.04]	0.03 [0.00,0.07]	0.03 [0.00,0.07]	0.05 [0.00, 0.10]	0.08 [0.01, 0.04]	0.09 [0.00, 0.10]	

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

attendance in Malawi because the study did not include this indicator. 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 in Zimbabwe for attendance. It is unclear why Kenya has higher ICC education values, especially for enrollment. The higher ICC values for education outcomes in Kenya will affect the ability to detect impacts for these indicators, all else equal, especially for primary school age children for whom the ICC is estimated as 0.23. Across the five studies, our estimates suggest that cluster-RCTs of education programs may have design effects as high as 7.2. However, this design effect may be larger than in typical education-focused cluster-RCTs with random assignment at the classroom level (or school-level with a sample from a small number of grades), because these studies often have a smaller numbers of students per classroom or school.

Fertility. Table 4 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 similar ICC values both within a country and across all five studies. The ICC values are quite low and stable with small confidence intervals. Thus, there is relatively 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 reduce the power of a study to detect effects by less than for other outcomes with higher ICC values. Our estimates suggest that the design effect of a cluster RCT for fertility outcomes is around 1.93.

Household Level Indicators

We estimate ICCs for poverty and food security, livestock and agricultural assets, agricultural production, consumption, and household living conditions at the household level.

Poverty and Food Security. Table 5 shows the ICC values for regularly measured poverty and food security indicators. Many of the cash transfer programs target beneficiaries based on poverty levels 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 expect to find fairly low ICCs for these indicators. In line with this hypothesis, the ICC values for these indicators are fairly low and consistent across countries and across indicators within the same country (see Table 5). Our estimates suggest a design effect of 5.09 for poverty and food security outcomes.

Table 4. ICCs for fertility.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Age of first pregnancy	0.01 [0.00,0.03]	0.02 [0.00,0.04]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.01 [0.00,0.02]	0.00 [0.00,0.00]	0.01 [0.00,0.02]	0.01 [0.00,0.02]	0.01 [0.00,0.02]	0.01 [0.00,0.02]	0.01 [0.00,0.02]	0.01 [0.00,0.02]
Number of children	0.01 [0.00,0.03]	0.03 [0.01,0.06]	0.01 [0.00,0.07]	0.02 [0.00,0.09]	0.02 [0.00,0.04]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.00 [0.00,0.01]	0.01 [0.00,0.01]	0.02 [0.00,0.01]

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Table 5. ICC values for household poverty and food security.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Below poverty line	0.09 [0.04,0.14]	0.10 [0.03,0.16]	0.02 [0.01,0.03]	0.01 [0.00,0.03]	0.01 [0.00,0.03]	0.00 [0.00,0.00]	0.01 [0.00,0.01]	0.00 [0.00, 0.00]	0.02 [0.00, 0.05]	0.02 [0.00, 0.04]	0.02 [0.00, 0.05]	0.03 [0.00, 0.04]
Severe poverty	0.07 [0.02,0.11]	0.03 [0.00,0.07]	0.06 [0.03,0.08]	0.06 [0.03,0.09]	0.04 [0.02,0.07]	0.03 [0.01,0.06]	0.04 [0.02,0.06]	0.03 [0.03, 0.03]	0.05 [0.01, 0.08]	0.04 [0.01, 0.06]	0.05 [0.01, 0.06]	0.04 [0.01, 0.10]
Food insecurity access scale			0.04 [0.02,0.06]	0.04 [0.04, 0.04]	0.16 [0.09,0.23]	0.15 [0.15, 0.15]	0.12 [0.07,0.17]	0.04 [0.02,0.06]				
Hunger scale			0.05 [0.03,0.07]	0.05 [0.05, 0.06]	0.11 [0.06,0.16]	0.12 [0.12, 0.12]	0.07 [0.03,0.10]	0.05 [0.03,0.08]				
Meals per day			0.07 [0.04,0.09]	0.08 [0.08, 0.08]	0.09 [0.04,0.13]	0.09 [0.09, 0.09]	0.05 [0.02,0.08]	0.08 [0.05,0.12]	0.07 [0.02, 0.12]	0.08 [0.03, 0.12]	0.07 [0.02, 0.12]	0.07 [0.03, 0.12]

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Livestock and Agricultural Assets. Table 6 shows the ICC values for the number of livestock owned and an index of agricultural assets. Most of the households in the five studies are rural, subsistence farming households. With the exception of Kenya, we find fairly consistent ICC values across countries for each indicator, resulting in fairly robust ICC values. However, we find more variation in ICC values across indicators. ICCs for the ownership of chickens and goats seem higher than for the ownership of pigs and cattle. Regardless, the ICC values for all of these indicators are below 0.1 and some are lower than 0.05. The ICC for agricultural assets appears to have more variation across countries, however, and also is higher than for most livestock categories, thus potentially affecting statistical power more than for the livestock indicators. Our estimates suggest that the design effect of a cluster RCT for agricultural livestock and asset outcomes is around 4.10.

Agricultural Production. Table 7 shows the ICC values for several agricultural production indicators. We find a large 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, however, so have fewer opportunities for comparisons. The ICC for crop sales appears to be the most consistent across countries, with the exception of Malawi, which has a much higher ICC for crop sales than in the other four countries. Our estimates suggest a design effect of 3.98 for agricultural outcomes.

Consumption. Table 8 shows consumption and expenditure ICC values for different consumption 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 ICCs for school enrollment observed in the education section. ICCs of health expenditures may be similar to ICCs of education expenditures, because some clusters have a clinic or school located near-by while others are further away, causing variation across clusters that is larger than the variation within a cluster. Across the disparate types of consumption and expenditure, our estimates suggest that a cluster RCT with 63 households per cluster typically has a design effect of 4.47.

Household Living Conditions. Table 9 shows the ICC values for household living conditions, specifically access to drinking water and access to a toilet or latrine. The values of the ICCs for these indicators have little consistency across studies with some of the highest values reported in Kenya, which may also be driven by urban data with different living conditions. These indicators also have some of the

Table 6. ICC values for livestock and agricultural assets.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Number of pigs	0.01 [0.00,0.02]	0.01 [0.00,0.03]	0.03 [0.01,0.04]	0.04 [0.02,0.06]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.01 [0.00, 0.01]	0.00 [0.00,0.01]	0.01 [0.00, 0.01]	0.00 [0.00,0.01]	0.01 [0.00,0.01]	0.01 [0.00,0.01]
Number of chickens	0.10 [0.05,0.16]	0.13 [0.05,0.20]	0.03 [0.01,0.05]	0.04 [0.02,0.06]	0.05 [0.02,0.08]	0.05 [0.02,0.08]	0.06 [0.03,0.09]	0.07 [0.03,0.10]	0.01 [0.00, 0.02]	0.01 [0.00, 0.02]	0.05 [0.00, 0.01]	0.05 [0.00, 0.01]
Number of goats	0.29 [0.18,0.41]	0.78 [0.69,0.87]	0.03 [0.02,0.05]	0.03 [0.01,0.05]	0.12 [0.06,0.19]	0.15 [0.07,0.23]	0.03 [0.01,0.04]	0.03 [0.01,0.05]	0.02 [0.00, 0.04]	0.02 [0.00, 0.04]	0.10 [0.00, 0.04]	0.20 [0.00, 0.04]
Number of cattle	0.10 [0.04,0.15]	0.14 [0.06,0.22]	0.03 [0.01,0.05]	0.02 [0.01,0.04]	0.01 [0.00,0.03]	0.02 [0.00,0.04]	0.00 [0.00,0.01]	0.01 [0.00,0.02]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.03 [0.01, 0.05]	0.05 [0.00, 0.06]
Index of agricultural productive assets			0.09 [0.06,0.12]	0.10 [0.06,0.13]	0.03 [0.01,0.05]	0.02 [0.00,0.04]	0.05 [0.03,0.08]	0.05 [0.02,0.08]	0.03 [0.01, 0.05]	0.04 [0.00, 0.06]	0.05 [0.00, 0.06]	0.05 [0.00, 0.06]

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Table 7. Agricultural production.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Crop sales	0.00	0.00	0.00	0.02	0.02	0.03	0.04	0.03	0.29	0.30	0.09	0.09
	[0.00,0.00]	[0.00,0.00]	[0.00,0.04]	[0.00,0.04]	[0.01,0.06]	[0.02,0.07]	[0.01,0.06]	[0.15, 0.43]	[0.16, 0.44]			
Value of consumption from own production	0.06	0.07	0.14	0.14	0.15	0.15	0.15	0.15	0.05	0.06	0.16	0.11
	[0.04,0.09]	[0.04,0.10]	[0.08,0.20]	[0.07,0.20]	[0.09,0.21]	[0.08,0.22]	[0.01, 0.08]	[0.02, 0.10]				
Agricultural productivity (\$/acre)			0.00	0.00	0.10	0.12	0.00	0.00	0.00	0.00	0.03	0.04
			[0.00,0.00]	[0.00,0.00]	[0.06,0.15]	[0.12, 0.12]	[0.00,0.00]	[0.00,0.01]				

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Table 8. Consumption ICC values.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Total consumption per capita	0.02 [0.00,0.03]	0.01 [0.00,0.03]	0.03 [0.02,0.05]	0.02 [0.01,0.04]	0.11 [0.06,0.16]	0.09 [0.04,0.13]	0.08 [0.04,0.12]	0.08 [0.03,0.12]	0.07 [0.02, 0.11]	0.06 [0.01,0.10]	0.06 [0.01,0.10]	0.05
Education expenditures per capita	0.13 [0.06,0.20]	0.11 [0.04,0.18]	0.05 [0.02,0.07]	0.04 [0.01,0.06]	0.00 [0.00,0.01]	0.00 [0.00,0.02]	0.02 [0.00,0.03]	0.02 [0.00,0.04]	0.03 [0.00, 0.05]	0.02 [0.00,0.04]	0.02 [0.00,0.04]	0.04
Health expenditures per capita	0.00 [0.00,0.01]	0.00 [0.00,0.01]	0.02 [0.01,0.04]	0.04 [0.01,0.06]	0.03 [0.01,0.05]	0.00 [0.00,0.02]	0.08 [0.04,0.12]	0.09 [0.04,0.13]	0.04 [0.01, .06]	0.04 [0.01,0.07]	0.04 [0.01,0.07]	0.03
Food expenditures per capita	0.06 [0.02,0.10]	0.07 [0.01,0.12]	0.05 [0.03,0.07]	0.05 [0.03,0.08]	0.12 [0.06,0.17]	0.10 [0.05,0.14]	0.07 [0.03,0.11]	0.07 [0.02,0.11]	0.06 [0.02, 0.09]	0.06 [0.02,0.10]	0.06 [0.02,0.10]	0.07

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

Table 9. ICC values for household living conditions.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi		Average ICC	
	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted
Access to drinking water			0.20 [0.15,0.26]	0.20 [0.15,0.26]	0.27 [0.18,0.36]	0.25 [0.16,0.35]	0.07 [0.03,0.10]	0.06 [0.02,0.10]	0.11 [0.04, 0.19]	0.11 [0.04, 0.19]	0.21 [0.04, 0.19]	0.20
Access to toilet/latrine	0.41 [0.28,0.55]	0.40 [0.26,0.54]	0.24 [0.18,0.31]	0.21 [0.15,0.28]	0.50 [0.39,0.61]	0.49 [0.38,0.60]	0.12 [0.06,0.17]	0.06 [0.02,0.11]	0.11 [0.04, 0.19]	0.11 [0.04, 0.19]	0.28 [0.04, 0.19]	0.25
Owms toilet on HH premises	0.07 [0.02,0.11]	0.04 [0.01,0.08]	0.23 [0.17,0.29]	0.22 [0.16,0.28]	0.42 [0.31,0.53]	0.41 [0.30,0.52]	0.21 [0.13,0.28]	0.19 [0.12,0.27]	0.07 [0.02, 0.12]	0.07 [0.02, 0.12]	0.19 [0.02, 0.12]	0.19
Purchased roof	0.68 [0.56,0.79]	0.63 [0.50,0.77]	0.21 [0.16,0.27]	0.21 [0.15,0.26]	0.02 [0.00,0.04]	0.01 [0.00,0.03]	0.04 [0.02,0.06]	0.03 [0.01,0.05]				0.24
Purchased cooking fuel	0.29 [0.18,0.41]	0.26 [0.15,0.37]	0.22 [0.15,0.28]	0.18 [0.12,0.24]	0.11 [0.06,0.17]	0.11 [0.05,0.16]	0.09 [0.04,0.13]	0.09 [0.04,0.14]			0.18 [0.04, 0.14]	0.16
Purchased lighting fuel	0.75 [0.65,0.85]	0.86 [0.79,0.93]	0.16 [0.12,0.21]	0.10 [0.06,0.14]	0.21 [0.13,0.29]	0.17 [0.09,0.24]	0.24 [0.16,0.32]	0.22 [0.14,0.31]			0.34 [0.14, 0.31]	0.34

Note: The numbers in brackets represent the 95% confidence interval for the true value of the ICC based on the standard error of the ICC estimate, the sample size, and a lower-bound of 0.00. Adjusted ICCs represent the share of the variance attributable to the cluster after accounting for household size and the gender, age, and primary school completion of the survey respondent.

highest ICC values reported in this paper, meaning that they will have larger implications for power of a cluster-randomized design. The ICC values for Malawi and the MCTG Zambia study are fairly consistent within each of these studies as they hover around 0.1, but the ICC values within each of the other studies vary widely across the indicators and are substantially higher, especially in Kenya. In Kenya, the ICC values are especially high because of relatively limited variation for some of the outcomes (i.e., 92% of the households had purchased housing fuel) and because of a more urban setting. The higher ICCs result in a very high design effect of up to 15.88, which showcases the importance of keeping cluster sizes smaller in studies targeting housing and other infrastructure outcomes when these outcomes are geographically clustered.

Conclusion

This paper presents ICC values for a range of indicators within domains commonly studied in development economics and for four countries in sub-Saharan Africa serving as a reference for researchers when designing clustered RCTs in the region. We show for which outcomes ICCs are stable or fluctuate by providing multiple indicators within a domain and similar indicators across countries. The paper does not only provide ICC values for these countries/contexts, but also enables researchers to determine when an ICC might generalize to a similar country in the region when it is not listed here. A main finding from the paper is that the ICC values we present are lower than 0.10 in most cases, meaning that studies lose less power to detect effects when moving from individual to clustered randomization designs in comparison with situations in which ICCs are higher than 0.10. ICCs for the indicators associated with consumption, education, fertility, and nutrition are all lower than 0.10 in each of the Southern African countries in our sample. ICC values of indicators associated with agriculture, food security, and livestock are generally lower than 0.10, but we find higher ICCs and design effects for some of the housing and other infrastructure indicators. Many of the ICCs we present are lower than 0.05 or even 0.01, however. For example, 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 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. All findings only change marginally after using covariate-adjusted ICCs, which generally result in slightly smaller ICCs (ranging from 0.01–0.03).

We find some 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. We propose that the inclusion of urban settings drives up ICCs across outcomes, with different values across urban and rural settings and higher ICCs in urban settings.² These differences show the importance of accounting for 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. The evaluations in this paper studied programs with very similar targeting mechanisms to identify program beneficiaries. The ICCs in this paper relate to predominantly rural populations of poor, food insecure, and labor constrained households. The ICCs might look different for the same indicators in the same countries if the studies targeted a different population. Therefore, it is important to consider the population served by a program in a particular evaluation before generalizing the ICCs from that evaluation to another study.

The number of impact evaluations in international development has increased significantly since 2009 (Brown et al. 2016). Meanwhile, 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 et al. (2017) show that in half of the field of economics' research areas, 90% or more of its studies are underpowered. In addition, Ioannidis (2005) shows that low power increases the likelihood of false positives. At the same time, Banerjee et al. (2015) highlight the importance of conducting ex-ante power calculations in development economics research. This paper can help researchers more accurately estimate their minimum detectable effect size and design well powered studies to avoid some of these problems of statistical power found in the literature.

In addition to providing more ICCs for other countries and domains, future research could also focus on estimating the other parameters that are relevant for power calculations in a wide range of settings. 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. 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.

Appendix

Annex A. Descriptive Tables

Annex A1. Descriptive Statistics for Child Nutrition.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Stunted (0–2 years)	0.28	0.45	0.17	0.38	0.31	0.46			0.34	0.47
Stunted (0–5 years)	0.33	0.47	0.27	0.44	0.34	0.47			0.45	0.50
Severely stunted (0–2 years)	0.12	0.33	0.05	0.22	0.11	0.32			0.15	0.36
Severely stunted (0–5 years)	0.14	0.35	0.09	0.29	0.12	0.33			0.20	0.40
Height-for-age z-score (0–24 months)	−0.81	2.11	−0.65	1.66	−1.31	1.63			−1.46	1.59
Height-for-age z-score (0–60 months)	−1.08	2.01	−1.17	1.48	−1.41	1.60			−1.78	1.59
Wasted (0–2 years)	0.09	0.29	0.03	0.16	0.09	0.29			0.08	0.27
Wasted (0–5 years)	0.08	0.27	0.02	0.15	0.06	0.24			0.04	0.20
Severely wasted (0–2 years)	0.02	0.16	0.00	0.06	0.02	0.15			0.02	0.14
Severely wasted (0–5 years)	0.03	0.16	0.01	0.07	0.02	0.13			0.01	0.10
Weight-for-height z-score (0–24 months)	0.03	1.61	0.01	1.05	−0.30	1.32			−0.19	1.26
Weight-for-height z-score (0–60 months)	−0.14	1.45	0.03	1.04	−0.18	1.23			0.10	1.15
Underweight (0–2 years)	0.19	0.39	0.08	0.27	0.18	0.38			0.18	0.38

(continued)

Annex A1. (continued)

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Underweight (0–5 years)	0.21	0.41	0.10	0.31	0.16	0.37			0.17	0.38
Severely underweight (0–2 years)	0.08	0.27	0.02	0.16	0.05	0.22			0.05	0.22
Severely underweight (0–5 years)	0.08	0.26	0.02	0.15	0.04	0.21			0.04	0.20
Weight-for-age z-score (0–24 months)	–0.59	1.81	–0.30	1.30	–0.87	1.33			–0.85	1.31
Weight-for-age z-score (0–60 months)	–0.89	1.56	–0.61	1.17	–0.92	1.23			–0.95	1.20

Annex A2. Descriptive Statistics for Child Nutrition.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Currently enrolled (7–14 years)	0.90	0.30	0.92	0.28	0.72	0.45	1.00	0.02	0.77	0.42
Days attended (7–14 years)	4.36	1.50	3.90	1.78	4.47	1.23	4.53	1.28		
Days attended (15–18 years)	3.82	1.97	2.17	2.36	4.49	1.26	4.62	1.20		
Full attendance in past week (7–14 years)	0.66	0.47	0.62	0.48	0.79	0.41	0.84	0.37	18.53	0.35
Full attendance in past week (15–18 years)	0.56	0.50	0.36	0.48	0.80	0.40	0.88	0.33	3.91	0.34

Annex A3. Descriptive Statistics for Fertility.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age of first pregnancy			19.14	3.52	20.58	4.36	26.08	6.72	18.53	2.96
Number of children			3.45	2.09	4.01	2.02	4.06	2.39	3.91	2.22

Annex A4. Descriptive Statistics for Household Poverty and Food Security.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Below poverty line	0.81	0.39	0.93	0.26	0.99	0.12	0.96	0.19	0.90	0.29
Severe poverty	0.53	0.50	0.69	0.46	0.95	0.23	0.89	0.31	0.73	0.44
Food insecurity access scale			13.46	6.21	15.15	5.68	14.67	5.54		
Hunger scale			1.16	1.56	2.95	1.67	2.56	1.87		
Meals per day			2.16	0.61	1.87	0.53	1.77	0.54	1.91	0.57

Annex A5. Descriptive Statistics for Livestock and Agricultural Assets.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of pigs	0.03	0.33	0.09	0.63	0.02	0.23			0.01	0.17
Number of chickens	3.93	6.99	3.76	4.98	1.94	3.88	2.62	4.64	0.60	1.94
Number of goats	2.18	8.15	1.86	3.26	0.05	0.45	0.36	1.44	0.25	1.01
Number of cattle	1.12	2.16	0.90	1.73	0.61	5.83	0.03	0.37	0.00	0.08
Index of agricultural productive assets			0.00	0.95	0.00	0.79	0.00	0.75	0.00	1.32

Annex A6. Descriptive Statistics for Agricultural Production.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Crop sales			56	156	83285	394860	213156	586280	27215	31161
Value of consumption from own production			22	23	180	252	409347	370674	998	1294
Agricultural productivity (\$/acre)					1.26	13	401989	339822	1.38	3.05

Annex A7. Descriptive Statistics for Consumption.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total consumption per capita	13832	26883	35	36	40	32	51346	45262	45059	34479
Education expenditures per capita	1112	2449	1	3	3	12	7436	26733	1504	4680
Health expenditures per capita	1434	25210	2	3	12	17	9464	13636	373	1064
Food expenditures per capita	9684	6137	22	21	30	27	40094	38067	129952	95815

Annex A8. Descriptive Statistics for Household Living Conditions.

Indicator	Kenya		Zimbabwe		Zambia (CGP)		Zambia (MCTG)		Malawi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Access to drinking water	0.36	0.48	0.57	0.50	0.22	0.41	0.23	0.42	0.09	0.29
Access to toilet/ latrine	0.44	0.50	0.24	0.43	0.49	0.50	0.93	0.26	0.74	0.44
Owens toilet on HH premises	0.08	0.28	0.36	0.48	0.44	0.50	0.79	0.41	0.61	0.49
Purchased roof	0.77	0.42	0.28	0.45	0.05	0.23	0.10	0.30		
Purchased cooking fuel	0.08	0.27	0.03	0.17	0.05	0.22	0.03	0.17		
Purchased lighting fuel	0.92	0.27	0.63	0.48	0.58	0.49	0.59	0.49		

Acknowledgments

The authors would like to thank Joshua Sennett, Frank Otchere, and John Downes for their key contributions to data analysis. This study would have not have been possible without the support of the UNICEF country offices that commissioned the program evaluations.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The article has not been published or submitted for publication in whole or part to any other publishing source. I certify that the underlying analysis is in compliance with the governing IRB standards at the American Institutes for Research.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Thomas de Hoop  <https://orcid.org/0000-0002-0033-1139>

Mitchell Morey  <https://orcid.org/0000-0002-1469-1872>

Notes

1. All survey tools are available from The Transfer Project portal: <https://transfer.cpc.unc.edu/instruments/>. The data used in this study are available for download or for request at the University of North Carolina's CPC Data Portal: <https://transfer.cpc.unc.edu/datasets/>. Baseline reports describing full descriptive results of each study

(including mean values and standard deviations) are available on The Transfer Project portal: <https://transfer.cpc.unc.edu/publications/>. Annex A reports means and standard deviations for all outcomes.

2. Across all outcomes, there is an average difference between urban and rural ICCs of 0.085 with 81% of outcomes having higher ICCs in urban settings compared to rural settings.

References

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35. <https://doi.org/10.1093/qje/qjac038>
- Angrist, J. D., & Jörn-Steffen, P. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *The Journal of Economic Perspectives*, *24*(2), 3–30. <https://doi.org/10.1257/jep.24.2.3>
- Ballard, T., Coates, J., Swindale, A., & Deitchler, M. (2011). *Household hunger scale: Indicator definition and measurement guide*. Washington, DC: Food and Nutrition Technical Assistance II Project. FHI 360.
- Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1–21. <https://doi.org/10.1257/app.20140287>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., EmmaRobinson, S. J., & Munaf, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Cameron, Drew, Mishra, Anijni, & Brown, Annette (2016). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, *8*(1), 1–21.
- Campbell, M. K., Grimshaw, J. M., & Elbourne, D. R. (2004). Intraclass correlation coefficients in cluster randomized trials: Empirical insights into how should they be reported. *BMC Medical Research Methodology*, *4*(1), 9. <https://doi.org/10.1186/1471-2288-4-9>
- Cohen, J., Erlbaum Associates (1977). *Statistical power analysis for the behavioral sciences rev.* Lawrence: Inc.
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, *4*(1), 3895–3962.
- Geyer, J., Davis, M., & Narayan, T. (2016). Intraclass correlation coefficients of household economic and agricultural outcomes in Mozambique. *Evaluation Review*, *40*(6), 526–545. <https://doi.org/10.1177/0193841x16659834>
- Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). 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(1), 876–883. <https://doi.org/10.1093/oxfordjournals.aje.a009904>
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis* 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>
- Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two-and three-level cluster-randomized experiments in education. *Evaluation Review*, 37(6), 445–489. <https://doi.org/10.1177/0193841x14529126>
- Heneveld, W., & Craig, H. (1996). *Schools count: World Bank project designs and the quality of primary education in Sub-Saharan Africa*. The World Bank.
- Ioannidis, J., & Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys*, 27(1), 997–1004. <https://doi.org/10.1111/joes.12032>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS med*, 2(1), Article124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 127(605), F236–F265. <https://doi.org/10.1111/eoj.12461>
- Isaakidis, P., & Ioannidis, J. P. (2003). Evaluation of cluster randomized controlled trials in sub-Saharan Africa. *American Journal of Epidemiology*, 158(9), 921–926. <https://doi.org/10.1093/aje/kwg232>
- Jacob, R., Zhu, P., & Bloom, H. (2010). New empirical evidence for the design of group randomized trials in education. *Journal of Research on Educational Effectiveness*, 3(1), 157–198.
- Kelcey, B., Shen, Z., & Spybrook, J. (2016). Intraclass correlation coefficients for designing cluster-randomized trials in sub-Saharan Africa education. *Evaluation Review*, 40(1), 500–525. <https://doi.org/10.1177/0193841x16660246>
- Kelcey, B., Spybrook, J., Phelps, G., Jones, N., & Zhang, J. (2017). Designing large-scale multisite and cluster-randomized studies of professional development. *The Journal of Experimental Education*, 85(3), 389–410. <https://doi.org/10.1080/00220973.2016.1220911>
- McKenzie, D. (2012). Tools of the trade: Intra-cluster correlations. *Impact evaluations*. World Bank.
- Murray, D. M., & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review*, 27(1), 79–103. <https://doi.org/10.1177/0193841x02239019>
- Shoukri, M. M., Donner, A., & El-Dali, A. (2013). Covariate-adjusted confidence interval for the intraclass correlation coefficient. *Contemporary Clinical Trials*, 36(1), 244–253. <https://doi.org/10.1016/j.cct.2013.07.003>
- Spybrook, J., & Kelcey, B. (2016). Introduction to three special issues on design parameter values for planning cluster randomized trials in the social sciences. *Evaluation Review*, 40(6), 491–499. <https://doi.org/10.1177/0193841x16685646>

- Ukoumunne, O. C., Gulliford, M. C., & Chinn, S. (2002). A note on the use of the variance inflation factor for determining sample size in cluster randomized trials. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(4), 479–484. <https://doi.org/10.1111/1467-9884.00332>
- Westine, C. D., Spybrook, J., & Taylor, J. A. (2013). An empirical investigation of variance design parameters for planning cluster-randomized trials of science achievement. *Evaluation Review*, 37(6), 490–519. <https://doi.org/10.1177/0193841x14531584>
- Wooldridge, & Jeffrey, M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- Zopluoglu, C. (2012). A cross-national comparison of intra-class correlation coefficient in educational achievement outcomes. *Eğitimde Ve Psikolojide Ölçme Ve Değerlendirme Dergisi*, 3(1), 242–278.