Recalling Extra Data: A Replication Study of Finding Missing Markets

Benjamin D. K. Wood and Michell Dong

Selected Paper prepared for presentation at the International Agricultural Trade Research Consortium's (IATRC's) 2014 Annual Meeting: Food, Resources and Conflict, December 7-9, 2014, San Diego, CA.

Copyright 2014 by Benjamin D.K.Wood and Michell Dong. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Benjamin D. K. Wood
Michell Dong

Recalling extra data

A replication study of
Finding missing markets

May 2015

Replication
Paper 5

Agriculture

International
Initiative for
Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The **3ie Replication Programme** also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's Replication Window with generous funding from an anonymous donor. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, Benjamin DK Wood at bwood@3ieimpact.org.

Suggested citation: Wood, BDK and Dong, M, 2015. *Recalling extra data: a replication study of Finding missing markets*, 3ie Replication Paper 5. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editor: Benjamin DK Wood
Assistant managing editor: Jennifer Ludwig
Production manager: Jennifer Ludwig
Layout Assistant: Kara Ingraham
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill
Printer: VIA Interactive

© International Initiative for Impact Evaluation (3ie), 2015

Recalling extra data: a replication study of *Finding missing markets*

Benjamin Douglas Kuflick Wood
International Initiative for Impact Evaluation

Michell Dong
World Bank

3ie Replication Paper 5
May 2015

Acknowledgments

We would like to thank Annette Brown, Howard White, Heather Lanthorn, Eric Djimeu, Lori Beaman and four anonymous external reviewers for helpful comments on this research. We would also like to thank Sebastian Insfran Moreno for his early-stage research assistance and Courtney Soderberg for sharing her statistician perspective on power calculations in this context. Finally, we extend our gratitude to Nara Ashraf, Xavier Giné and Dean Karlan, the authors of the original study, for providing us with the original data and survey instruments that made this replication study possible.

Abstract

We reexamine some of the strongest evidence supporting cash-crop-based development strategies by replicating Nara Ashraf, Xavier Giné and Dean Karlan's 'Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya', published in the *American Journal of Agricultural Economics*. The original evaluation, of an agricultural export crop promotion intervention in Kenya, is one of the few impact evaluations exploring how agricultural commercialisation affects household outcomes. Our attempt to independently reconstruct the evaluation using the existing raw data finds the original results generally robust to replication, albeit with much lower coefficients on some of the main outcomes of interest. We explore the evaluation's theory of change, focusing on the result that first-time export crop adopters benefit more from agricultural commercialisation than agricultural households that were already producing export crops. We also examine questions around adequate power requirements and potential recall and or courtesy bias within the analysis. Reproducing these original results is relevant both to encourage policymakers to use this evidence and to highlight knowledge gaps for future research.

Keywords: replication, Kenya, cash crops, adoption, heterogeneous impacts

Contents

Acknowledgments ..... **ii**
Abstract ..... **iii**
List of figures and tables ..... **v**
Abbreviations and acronyms ..... **vi**
1. Introduction **1**
1.1 Reviewing the intervention 2
1.2 Introducing the replication 2
2. The pure replication **3**
2.1 The data 3
2.2 Assignment to treatment, table 1 4
2.3 Reproducing the summary statistics 7
2.4 Reproducing the main results 9
2.5 Exploring heterogeneous impacts 11
2.6 Pure replication challenges 12
2.7 Conversion factors 13
2.8 Pure replication conclusions 14
3. Measurement and estimation analysis: power **14**
4. Theory of change analysis: reanalysis and alternative heterogeneity analysis **18**
5. Limitations **23**
6. Conclusions **23**
Appendix A: Tables **25**
References **35**

List of figures and tables

Figure 1 Baseline data distribution of land devoted to cash crops, by dataset 16
Figure 2 Previous adopters of export crops, by dataset 19
Figure 3 Previously sold to market, original dataset 20

Table 1A: Replication results of pre-intervention balance with original group treatment assignment, AGK table 1 reproduction 4
Table 1B: Replication results of pre-intervention balance with 'follow-up' group treatment assignment, AGK table 1 reproduction 5
Table 1C: Pre-intervention balance with 'updated' group treatment assignment, AGK table 1 reproduction 6
Table 2A: Updated baseline summary statistics, AGK table 2 partial reproduction 8
Table 2B: Comparison of AGK and replication baseline sample sizes 9
Table 3: Impacts of DrumNet: ITT OLS, AGK table 4 partial reproduction 10
Table 4A: Impacts of DrumNet (prior exporters versus new adopters): ITT OLS, AGK table 6 partial reproduction 11
Table 4B: Impacts of DrumNet (prior exporters versus new adopters): ITT OLS, AGK table 6 partial reproduction 12
Table 5: Respondents reporting in standard units of measure 14
Table 6: Updated baseline DrumNet summary statistics by dataset: replication results 15
Table 7: Impacts of DrumNet: ITT OLS by dataset 17
Table 8: Post-intervention power calculations, baseline by dataset 18
Table 9A: Impacts of DrumNet, ITT OLS heterogeneous outcomes, by baseline characteristics and dataset 21
Table 9B: Impacts of DrumNet, ITT OLS heterogeneous outcomes, by baseline characteristics and dataset 22

Appendix Table 1: Variable definition and construction 25
Appendix Table 2A: Number of SGHS, replication and original results 27
Appendix Table 2B: SGHS and Observations by Assignment Rule 27
Appendix Table 3: Number of observations using 'updated' Assignment, AGK appendix table 2 reproduction 28
Appendix Table 4: Baseline summary statistics using 'updated' 29
Appendix Table 5: Intervention participation factors, AGK table 3 reproduction 30
Appendix Table 6: OLS intervention impacts, AGK table 4 reproduction 31
Appendix Table 7: OLS intervention impact by treatment arm, AGK table 5 reproduction 31
Appendix Table 8A: Intervention impact by production OLS, AGK table 6 reproduction 32
Appendix Table 8B: Intervention impact by production OLS, AGK table 6 reproduction 32
Appendix Table 9A: Intervention impact by production history and arm OLS, AGK table 7 reproduction 33
Appendix Table 9B: Intervention impact by production history and arm OLS, AGK table 7 reproduction 33
Appendix Table 10: Intervention impact on prices OLS, AGK table 8 reproduction 34

Abbreviations and acronyms

3ie	International Initiative for Impact Evaluation
AGK	Nara Ashraf, Xavier Giné and Dean Karlan
Extra 500	Additional 500 households introduced to the survey after the baseline
ICC	Intra-cluster correlation coefficient
ITT	Intent-to-treat
OLS	Ordinary least squares
SHG	Self-help group

1. Introduction

Our replication study seeks to reconstruct the findings in 'Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya' (Ashraf, Giné and Karlan 2009a) using the original raw data. This innovative study in Kenya by Nara Ashraf, Xavier Giné and Dean Karlan (henceforth referred to as AGK) captures the much promoted yet under-researched concept of agricultural commercialisation and diversification through export-oriented crop promotion in the developing world.[1] Specifically, AGK evaluate the household-level income effects of efforts to encourage Kenyan farmers to adopt new export crops. Previous explorations of this topic generally rely on correlations or instrumental variable estimation strategies and only look at one aspect of commercialisation. AGK's experimental design and package of commercialisation tools allow for a cleaner causal interpretation of the intervention's impact.

Development professionals generally present agricultural commercialisation as a path to poverty alleviation. Previous research shows households that adopt cash crops in which they maintain a comparative advantage to consistently reach higher income levels.[2] But many of these studies use village, region or even nationally representative cross-sectional data, which make capturing the adoption process and assigning intervention attribution problematic.[3] AGK's experimental design increases the importance of their findings and helps justify our study.

Agricultural commercialisation is a highly policy-relevant topic. International agencies continue to draw attention to agribusiness development opportunities because of both their ability to support rural households and the possibility of averting future food crises. The Byerlee *et al.* (2013) highlights the importance of unlocking the potential of agricultural commercialisation, while describing some of the market failures that prevented previous efforts from being fully realised. Fan *et al.* (2015) explain how properly incentivised agricultural commercialisation supports economic development.

Transitioning from consumption farming to income farming represents a fundamental shift in the development process. Many developing world populations concentrate in rural areas and depend on agricultural production. Agricultural commercialisation often presents opportunities for rural households to exploit their comparative advantage by growing labour-intensive agricultural commodities. Generally, this commercialisation is advocated as a means of agricultural diversification into high-value crops, while allowing for the purchasing of required food. But to reap the benefits of commercialisation, a number of potential market failures, from unfamiliar export crops to capital constraints, must be overcome.

[1] See Byerlee *et al.* (2013) and Kherallah (2002) for examples of export crop promotion efforts.
[2] See Carletto *et al.* (2011) for more information on the micro-level welfare gains typically associated with nontraditional crop adoption in the developing world. Wiggins *et al.* (2011) provide a general overview of agricultural commercialisation for smallholders in Africa. Strasberg *et al.* (1999) provide an early overview of Kenyan agricultural commercialisation possibilities.
[3] Obare (2000), in an exception to the trend, uses a small district-level survey to discuss some of the impediments to agricultural commercialisation in a Kenyan context. While we find this research helpful for contextualising the research area, it does not answer the general effectiveness questions surrounding agricultural commercialisation.

1.1 Reviewing the intervention

The intervention AGK evaluate was designed to increase agricultural commercialisation by providing a package of extension and marketing services to treatment smallholder farmers. After identifying high-value international crops with strong local growth potential, the intervention targeted smallholder farmers with a package of services including current price information for those crops, linkages to a transportation supply chain and, for one treatment arm, linkages with commercial banks. These services were provided to overcome general constraints faced by smallholder farmers engaged in horticulture.

DrumNet, the project implementers, specifically targeted smallholder farmers through existing self-help groups (SHGs). They based intervention eligibility on agricultural household membership in a SHG registered with the Kenyan government; SHG interest in growing export crops; household access to irrigated land; and household ability to make a minimum payment of about US$10. DrumNet gave intervention recipients a month-long course in Good Agricultural Practices and instructions on opening a local bank account. They also gave households in the credit arm access to microcredit, with a minimum deposit required of those households to guarantee the loans they received.

This paper is based on a plausible causal chain. Farmers gain knowledge and skills in growing high-value export crops. The adoption of these crops, along with a reduction in transportation costs, leads to increases in household income. In the long run, it is hoped that this increase in household income will lead to general welfare benefits.[4] The evaluation also examines possible heterogeneous impacts based on households that previously produced export crops versus first-time adopters, as first-time adopters may have more capacity for increasing their household income.

1.2 Introducing the replication

The original study uses a randomised evaluation framework to test the impact of providing a package of agricultural commercialisation services to SHGs. AGK conclude that the intervention only increases the household income of first-time export crop adopters. Possible recall bias and power constraints may influence the strength of the original results. Following our posted replication plan, our study aims to better understand the robustness of the existing agricultural commercialisation evidence and highlight the effectiveness of this package of agricultural commercialisation interventions. [5]

The paper follows Brown *et al.* (2014) by including three main sections: pure replication, measurement and estimation analysis and theory of change analysis. In the pure replication section we explain the data, methods and assumptions we use to reevaluate the intervention. Our measurement and estimation analysis examines the different datasets used in the original analysis and includes a power analysis of the original study's sample size. Our theory of change analysis briefly explores alternative methods of analysing heterogeneous impacts by considering whether the types of crop being planted or the method of entering the commercialisation market is of more relevance to this intervention.

[4] As noted in the epilogue of the original paper, changes in Europe's food import certification system prevent the evaluation, and us, from estimating the long-term effects of this intervention.
[5] Our replication plan is available at http://www.3ieimpact.org/evaluation/impact-evaluation-replication-programme/replication-finding-missing-markets-and-disturbing-epilogue-evid/. When discussing the replication study findings, we note when and why we deviated from our plan.

2. The pure replication

Our pure replication uses the raw data to reassess the intervention. With that objective in mind, we reconstruct the original paper using only the raw data and the publication, with the survey instruments as a pseudo-codebook. We see major strengths and weaknesses associated with this replication approach. On the one hand, by coding the entire paper, we ensure our results to be independent of the original research findings. On the other hand, that independence comes at the cost of not having access to the enumerators to understand issues encountered in the field, not guaranteeing a similar approach to outliers in the data, and not necessarily following the same path as AGK on a host of other decisions that occur throughout an evaluation.

2.1 The data

The original study includes three datasets: the baseline, the follow-up, and the 'extra 500', which we explain in more detail below.[6] AGK shared with us the survey instruments, the raw data, and two codebooks that generally describe the data but do not contain detailed explanations of how they generated the study variables.[7] AGK pool these datasets together to generate most of their research findings.

The baseline and follow-up data are fairly standard datasets. They include information on the more than 700 households interviewed initially in the spring of 2004 and again a year later. The baseline data are broken into 15 individual datasets, each accounting for 1 to 2 pages of the survey instrument. The follow-up data are structured in a similar manner, in 12 datasets. These data cover a wide range of topics related to individual, household and regional issues related to agricultural production and the networks that may influence household decision-making.

The extra 500 dataset contains different types of data. Budget constraints forced AGK to initially interview a relatively small sample. To address possibly inadequate sample sizes AGK expanded their sample in the middle of the evaluation.[8] As they did not have baseline information for these new households, AGK collected the information retrospectively through recall questions at the time of the follow-up. These data are split into 13 datasets, each accounting for 1 to 2 pages of the survey instrument. AGK provided us with some additional data beyond with these three main datasets, including administrative and networking data. The administrative data gave us key information to help identify households across the different data collection rounds.

Although AGK provided us with the original data and replied to some of our questions about the data analysis, we always planned to independently reconstruct the original evaluation. While we try to avoid working with the original Stata .do files, we use some of AGK's code to

[6] A savvy reviewer noted that the 'extra 500' data do not actually contain 500 household observations. We generally follow AGK's naming convention in this regard. An additional 500 households were interviewed but AGK only include a subsample of them in the analysis due to intervention eligibility.
[7] AGK could not locate the formal follow-up survey instrument, but the composite version they provided us is relatively accurate in relation to the dataset.
[8] AGK explain their data collection process in Ashraf, Giné and Karlan (2009a) in the text on page 976 and in footnote 8 of their paper. We did not focus on recall in our replication plan, as we did not entirely understand the data collection timeline at that point in the replication process. See Beegle, Carletto and Himelein (2012) for a greater discussion of issues with recall bias in agricultural data.

reconstruct their risk preference variables, as AGK base these variables on hypothetical questions not included in the survey instrument. Also, due to some difficulties with matching households across rounds of the surveys, we eventually followed some of AGK's cleaning methods. We strive to reconstruct the datasets as accurately as possible.

2.2 Assignment to treatment, table 1

We begin our pure replication by reproducing the SHG-level balance statistics reported in the first table of the original paper. These summary statistics results compare the pre-intervention SHG characteristics between control and treatment groups, with the treatment groups separated into credit and no-credit arms. AGK (2009a) note that they randomly assigned the 36 SHGs into 3 groups of 12 SHGs (p. 976) before implementing the intervention.[9] Because the intervention follows random assignment, we expect the three experimental groups to share similar baseline characteristics.

Table 1A: Replication results of pre-intervention balance with original group treatment assignment, AGK table 1 reproduction

	Number of Observations	Means			p-Value	Means		p-Value of F-test
		All	Control	Combined Treatment		Credit	No Credit	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current number of members	36	28.7	31.4	27.3	0.51	25.3	29.3	0.70
		(17.5)	(19.6)	(16.6)		(10.9)	(21.2)	
Age of SHG (months)	36	4.77	4.99	4.66	0.85	3.61	5.71	0.58
		(4.89)	(3.9)	(5.39)		(2.23)	(7.3)	
SHG has social activities (1=yes)	36	0.53	0.75	0.42	0.06*	0.42	0.42	0.18
		(0.51)	(0.45)	(0.5)		(0.51)	(0.51)	
Fee contribution to the SHG per member	36	103	87.5	111	0.54	112	109	0.83
		(106)	(56.9)	(124)		(133)	(120)	
SHG has an account in the bank (1=yes)	36	0.64	0.67	0.63	0.81	0.58	0.67	0.90
		(0.49)	(0.49)	(0.49)		(0.51)	(0.5)	
Main road paved (1=yes)	36	0.86	1.00	0.79	0.09*	0.75	0.83	**0.21**
		(0.35)	(0)	(0.41)		(0.45)	(0.4)	
Distance to main market (km)	36	5.82	5.08	6.19	0.39	5.79	6.58	0.61
		(3.6)	(3.2)	(3.79)		(2.92)	(4.6)	
Time to the main market (minutes)	36	41.5	22.5	51.0	0.09*	68.8	33.3	**0.04****
		(47.1)	(16)	(54.6)		(70.3)	(24.5)	

Note: Statistically significant differences between the original findings and replication results set in bold within the table. Throughout the tables, results statistical significance is noted as * p < 0.01, ** p < 0.05, *** p < 0.01.

The first table comprises eight variables: *current number of members*, *age of SHG*, *existence of social activities*, *fee contributions to the SHG per member*, *existence of SHG bank account*, *paved main road*, *distance to the main market* and *time to the main market*.

[9] Targeting treatment at the SHG level allows for greater effectiveness because the intervention reaches a larger group of people with a lower cost. Focusing on SHGs also alleviates some unobserved-characteristic concerns among smallholder farmers, as the farmers may have similar motivation levels, in terms of willingness to adopt alternative crops. Future researchers should note the SHG intervention participation requirement when they consider the generalisability of the findings.

Due to the unavailability of raw SHG-level data, we use AGK's pre-constructed variables to analyse the pre-intervention balance between the different SHGs.

Using the treatment status from the randomisation data, we calculate the means and standard errors of the group characteristics by group treatment status. Because we use the pre-constructed SHG variables, we unsurprisingly find the exact same results as the original paper for the means and standard errors of the control and combined treatment groups in columns (1) to (4) of AGK's table 1 in our table 1A. The exactness of our replication results stops there, with all of the values of the means and the standard errors for the individual treatment arms differing in columns (5) to (7).

The differences in our pre-intervention summary statistic finding leads us to explore alternative possible assignments to treatment. The follow-up data include a *grptype* variable that appears to indicate the treatment status for each household. We compare the treatment assignment defined in the follow-up dataset by the *grptype* variable to that in the randomisation dataset and discover four SHGs with conflicting treatment statuses. According to these 'follow-up' treatment assignments, there are 14 SHGs in the control group, 11 in the credit group and 11 in the no-credit group.

Table 1B: Replication results of pre-intervention balance with 'follow-up' group treatment assignment, AGK table 1 reproduction

		Means			p-Value	Means		p-Value of F-test
	Number of Observations	All	Control	Combined Treatment		Credit	No Credit	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current number of members	36	28.7	29.8	28.0	0.78	25.3	31.0	0.72
		(17.5)	(19.7)	(16.5)		(10.9)	(21.3)	
Age of SHG (months)	36	4.77	6.51	3.78	0.11	3.61	3.97	0.28
		(4.89)	(6.63)	(3.35)		(2.23)	(4.37)	
SHG has social activities (1=yes)	36	0.53	0.77	0.39	**0.03***	0.42	0.36	**0.09**
		(0.51)	(0.44)	(0.5)		(0.51)	(0.5)	
Fee contribution to the SHG per member	36	103	88.5	111	0.55	112	110	0.84
		(106)	(54.6)	(127)		(133)	(126)	
SHG has an account in the bank (1=yes)	36	0.64	0.69	0.61	0.63	0.58	0.64	0.86
		(0.49)	(0.48)	(0.5)		(0.51)	(0.5)	
Main road paved (1=yes)	36	0.86	0.92	0.83	**0.43**	0.75	0.91	**0.42**
		(0.35)	(0)	(0.39)		(0.45)	(0.3)	
Distance to main market (km)	36	5.82	4.77	6.41	0.19	5.79	7.09	0.30
		(3.6)	(3.3)	(3.71)		(2.92)	(4.46)	
Time to the main market (minutes)	36	41.5	22.3	52.4	0.06*	68.8	34.5	**0.04****
		(47.1)	(15)	(55.4)		(70.3)	(25.3)	

Note: Statistically significant differences between the original findings and replication results set in bold within the table.

The follow-up alternative treatment assignment results in table 1B differ in a few ways from the previous and the original findings. For example, control group SHG members tend to be older than their counterparts, and combined treatment group SHG members required more

time on average to travel to the main market than SHG members in the control group. Other differences between these control and treatment groups identified in previous tables are not present in these results.

The results in table 1B continue to differ from the original results, encouraging us to consider alternative treatment assignment possibilities. Working with the mean values and the data, we determine that shifting one SHG from the no-credit to the credit group creates an exact replicate of AGK's first table. However, the experimental SHGs groups are imbalanced, with 12 in the control group, 13 in the credit group and 11 in the no-credit group. Table 1C presents the balance statistics using these 'updated' treatment assignments. This alteration does not affect columns (1) to (4), because the first half of the table combines treatment SHGs into one column.

Table 1C exactly replicates AGK's first table, including their treatment arm post-estimation tests on each baseline characteristic. Consistent with the findings from table 1 of the original paper, SHGs in the credit arm in table 1C tend to be worse off in terms of infrastructure and remoteness to the market than those in the control and no-credit groups.[10]

Table 1C: Pre-intervention balance with 'updated' group treatment assignment, AGK table 1 reproduction

		Means				Means		
	Number of Observations	All	Control	Combined Treatment	p-Value	Credit	No Credit	p-Value of F-test
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current number of members	36	28.7	31.4	27.3	0.51	24.2	31.0	0.52
		(17.5)	(19.6)	(16.6)		(11.3)	(21.3)	
Age of SHG (months)	36	4.77	4.99	4.66	0.85	5.24	3.97	0.81
		(4.89)	(3.9)	(5.39)		(6.24)	(4.37)	
SHG has social activities (1=yes)	36	0.53	0.75	0.42	0.06*	0.46	0.36	0.16
		(0.51)	(0.45)	(0.5)		(0.52)	(0.5)	
Fee contribution to the SHG per member	36	103	87.5	111	0.54	111	110	0.83
		(106)	(56.9)	(124)		(128)	(126)	
SHG has an account in the bank (1=yes)	36	0.64	0.67	0.63	0.81	0.62	0.64	0.97
		(0.49)	(0.49)	(0.49)		(0.51)	(0.5)	
Main road paved (1=yes)	36	0.86	1.00	0.79	0.09*	0.69	0.91	0.07*
		(0.35)	(0)	(0.41)		(0.48)	(0.3)	
Distance to main market (km)	36	5.82	5.08	6.19	0.39	5.42	7.09	0.37
		(3.6)	(3.2)	(3.79)		(3.09)	(4.46)	
Time to the main market (minutes)	36	41.5	22.5	51.0	0.09*	65.0	34.5	0.06*
		(47.1)	(16)	(54.6)		(68.6)	(25.3)	

Note: This reproduction of AGK table 1 is identical to the original.

[10] We keep the SHG that switches treatment arm status within our balance analysis, although no members from that SHG remain in the follow up surveys after data cleaning. We also follow AGK's lead in referring to infrastructure in reference to the roads variable being discussed in tables 1A-1C.

As treatment status influences all of the evaluation results, we cross-reference our results with the original .do files. It appears that the SHGs used in AGK table 1 are the same as our updated table 1C, meaning that there are 12 control SHGs, 13 credit SHGs and 11 no-credit SHGs.[11] We use the updated treatment assignment for the remainder of our replication study, making the treatment assignment question moot because we combine all treatment SHGs into one group.

2.3 Reproducing the summary statistics

The baseline summary statistics AGK present in their second table generally demonstrate the effectiveness of their randomisation. They find the different treatment assignments to be very comparable, with some minor statistically significant differences between the treatment and control households at the time of the baseline survey in three variables: *loan from a formal institution, uses machinery and/or animal force* and *total spent in marketing* (transportation costs). As all of these factors plausibly relate to the cash crop adoption decision, AGK control for them in the original paper's analysis, and we control for them in the replication study.

As our study focuses on the intent-to-treat (ITT) analysis for the combined treatment arms, we examine the baseline summary statistic differences between the combined treatment households and the control households. Thus, in table 2A we present only the combined treatment group baseline summary statistics. We make this decision mostly due to difficulties we encounter with assigning within treatment group status.[12]

There are a few other notable differences between the replication study and the original summary statistics in table 2A. On average, we find control households statistically significantly more likely to be younger, to be a member of a SHG for longer and to have more total household income than their counterparts. Some of the differences relate directly to cash crop adoption, with control households being more likely to grow export crops and devote more of their land to cash crop production.

The baseline statistically significant differences between the control and treatment households in table 2A appear to bias the impact evaluation results downwards. We find households in control SHGs to be more likely to live near a paved road and require less time to market than the treatment groups in table 1. These same households are more likely to grow export crops in table 2A. While we control for these differences in the ITT analysis, these factors suggest the analysis may understate the influence of the intervention evaluation results.

In the context of comparing general treatment and control households, we find unexpected differences in the balance between our baseline sample statistics. We record a larger number of statistically significant different t-statistic estimates in our replication sample in table 2A. Although our replication sample differs from the original evaluation, we believe our efforts to reproduce the evaluation follow the same general path as the original study.

[11] We contacted AGK about our assignment concerns and they said we had discovered an 'error' in their .do file and that the assignment was equal between the control and two treatment arms.
[12] The treatment arms differ only in access to credit. AGK did not find any significant differences between the two arms in their main analysis. Copies of all tables with alternative treatment assignments are available upon request. Appendix table 2B includes a breakdown of the numbers of observations by treatment assignment rule. The replication study's version of table 2A is presented in its entirety in appendix table 3.

Table 2A: Updated baseline summary statistics, AGK table 2 partial reproduction

	Means			p-Value on t-Test of Difference (2) and (3)	AGK Original Means			AGK's p-Value on t-Test of Difference (6) and (7)
	All	Control	Treatment		All	Control	Treatment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Member								
Age of member	42.32	40.20	43.34	0.00***	41.2	39.3	42.2	0.17
	(12.62)	(12.46)	(12.57)		(12.2)	(11.9)	(12.2)	
Literacy	0.91	0.91	0.91	0.91	0.90	0.89	0.90	0.79
	(0.29)	(0.29)	(0.29)		(0.30)	(0.30)	(0.29)	
Risk tolerance	0.37	0.37	0.37	0.85	0.38	0.39	0.38	0.89
	(0.42)	(0.42)	(0.42)		(0.42)	(0.42)	(0.42)	
Months as member in SHG	53.13	58.33	50.20	0.01***	52.51	57.2	49.8	0.51
	(40.07)	(44.68)	(36.96)		(39.7)	(44.4)	(36.5)	
(i) Member of SHG is an officer (1=yes)	0.14	0.14	0.14	0.75	0.16	0.16	0.16	0.92
	(0.35)	(0.34)	(0.35)		(0.37)	(0.36)	(0.37)	
Deposit in a formal bank (1=yes)	0.70	0.73	0.69	0.25	0.69	0.70	0.69	0.77
	(0.46)	(0.45)	(0.46)		(0.46)	(0.46)	(0.46)	
Loan from formal institution (1=yes)	0.04	0.06	0.03	0.05**	0.04	0.06	0.03	0.03**
	(0.19)	(0.23)	(0.17)		(0.19)	(0.23)	(0.17)	
Logarithm of total annual household income	3.53	3.63	3.47	0.10*	3.49	3.59	3.44	0.30
	(1.17)	(1.16)	(1.18)		(1.20)	(1.19)	(1.20)	
Number of household members	4.59	4.52	4.63	0.50	4.59	4.55	4.61	0.79
	(2.09)	(2.10)	(2.08)		(2.09)	(2.12)	(2.08)	
Land								
Harvest yield per acre (in Ksh 100,000)	1.02	0.80	1.15	0.33	0.29	0.33	0.27	0.30
	(4.63)	(6.02)	(3.62)		(0.62)	(0.65)	(0.60)	
Proportion of land that is irrigated	0.37	0.37	0.36	0.84	0.40	0.39	0.40	0.87
	(0.31)	(0.29)	(0.32)		(0.31)	(0.29)	(0.32)	
Total landholdings	2.13	2.20	2.09	0.36	1.80	1.90	1.75	0.56
(Acres)	(1.93)	(1.91)	(1.93)		(2.05)	(2.36)	(1.89)	
Proportion of land devoted to cash crops	0.53	0.55	0.51	0.03**	0.58	0.59	0.57	0.54
	(0.28)	(0.28)	(0.28)		(0.25)	(0.24)	(0.26)	
Production								
Grows export crop	0.48	0.58	0.44	0.00***	0.46	0.55	0.41	0.15
(1=yes)	(0.50)	(0.49)	(0.50)		(0.50)	(0.50)	(0.49)	
Sells to market	0.34	0.36	0.33	0.53	0.39	0.41	0.38	0.54
(1=yes)	(0.47)	(0.48)	(0.47)		(0.49)	(0.49)	(0.49)	
Used hired labour	0.32	0.32	0.32	0.88	0..34	0.34	0.34	0.99
(1=yes)	(0.47)	(0.47)	(0.47)		(0.45)	(0.44)	(0.46)	
Uses machinery and/or animal force (1=yes)	0.05	0.07	0.03	0.00***	0.06	0.09	0.04	0.06*
	(0.21)	(0.26)	(0.18)		(0.23)	(0.28)	(0.19)	
Value of harvested produce	201.21	126.06	241.01	0.11	44.27	48.1	42.1	0.37
	(911.6	(1228.56)	(685.55)		(72.7)	(73.1)	(72.6)	
Production of French beans (in 1,000 kg)	0.75	0.54	0.85	0.40	3.40	2.89	3.65	0.61
	(5.55)	(1.20)	(6.66)		(14.3)	(13.1)	(14.9)	
Production of baby corn (in kg)	4.10	10.70	1.06	0.13	13.3	21.0	9.48	0.34
	(94.60)	(167.01)	(14.95)		(114.1)	(162.1)	(80.6)	
Total spent in marketing	1.30	0.88	1.53	0.16	1.00	0.36	1.36	0.06*
(in Ksh 1,000)	(5.95)	(3.64)	(6.91)		(8.18)	(2.13)	(10.1)	
Use of inputs	0.98	0.98	0.97	0.21	0.95	0.95	0.95	0.89
	(0.15)	(0.12)	(0.17)		(0.23)	(0.22)	(0.23)	

Note: Columns 1─4 report our results, and columns 5─8 report AGK's findings, and appendix table 3 records the number of households in our replication study for each of these variables.

Overall, the original paper and the replication study report very similar baseline sample sizes for each variable. Even still, our replication samples differ in the number of household observations, which influences all of our subsequent results. As shown in table 2B, small differences exist between the number of household baseline observations in the original and the replication, which become a bit larger when looking at the number of households in the specific treatment arms.

Table 2B: Comparison of AGK and replication baseline sample sizes

	New	Original	New	Original	New	Original	New	Original	New	Original
	All	All	Control	Control	Treatment	Treatment	Credit	Credit	No Credit	No Credit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Member										
Age of member	1110	1117	361	367	749	750	397	373	352	377
Literacy	1110	1117	361	367	749	750	397	373	352	377
Risk tolerance	726	726	261	263	465	463	217	216	248	247
Months as member in SHG	726	726	261	263	465	463	217	216	248	247
Member of SHG is an officer (1=yes)	1110	1117	361	367	749	750	397	373	352	377
Deposit in a formal bank (1=yes)	717	725	254	263	463	462	215	215	248	247
Loan from formal institution (1=yes)	726	726	261	263	465	463	217	216	248	247
Logarithm of total annual household income	707	713	255	259	452	454	214	215	238	239
Number of household members	722	726	260	263	462	463	215	216	247	247
Land										
Harvest yield per acre (in Ksh 100,000)	726	726	261	263	465	463	217	216	248	247
Proportion of irrigated land	1110	1117	361	367	749	750	397	373	352	377
Total landholdings (Acres)	1110	1117	361	367	749	750	397	373	352	377
Proportion of land devoted to cash crops	1037	990	327	302	710	688	375	344	335	344
Production										
Grows export crop (1=yes)	1037	1052	327	334	710	718	375	355	335	363
Sells to market (1=yes)	726	726	261	263	465	463	217	216	248	247
Used hired labour (1=yes)	1110	1117	361	367	749	750	397	373	352	377
Uses machinery and/or animal force (1=yes)	1110	1117	361	367	749	750	397	373	352	377
Value of harvested produce (in Ksh 1,000)	699	699	242	257	457	442	212	208	245	234
Production of French beans (in 1,000 kg)	1037	1051	327	334	710	717	375	355	335	362
Production of baby corn (in kg)	1037	1051	327	334	710	717	375	355	335	362
Total spent in marketing (in Ksh 1,000)	726	722	261	263	465	459	217	213	248	246
Use of inputs	1037	1032	327	317	710	715	375	354	335	361

Note: Columns 1, 3, 5, 7 and 9 report our observations, and columns 2, 4, 6, 8 and 10 report AGK's observations as recorded in their Appendix Table 2.

2.4 Reproducing the main results

Although AGK report a number of results, they show their key findings on the effectiveness of the agricultural commercialisation intervention mainly in their fourth table.[13] In this difference-in-difference table they present results from controlling for SHG fixed effects and some of the variables highlighted in their balance tables. The table includes 10 regressions,

[13] We have reproduced our version of the determinants of DrumNet participation outlined in AGK's table 3. As this section is not central to their argument, we present those results in appendix table 5. We made a similar determination about tables 5, 7 and 8 from the original publication, which are appendix tables 7, 9A, 9B and 10 in our paper.

each of which examines a dependent variable. These regressions show the impact of the DrumNet intervention on a number of outcomes of interest, including adoption of the export crops recommended by the intervention and changes in the log of household income.

The fourth AGK table includes both ITT and treatment-on-the-treated instrumental variable results. We reproduce only the ITT results, as AGK express strong validity concerns with their instrumental variable approach and present their ITT estimates as their preferred results. These ordinary least squares (OLS) regression findings are generally consistent across both of the estimation types in the original paper.

AGK examine the causal relationship between the DrumNet intervention and an increase in household income. This ultimate outcome of interest first requires households to grow export crops and devote more land to cash crop production. As DrumNet targeted baby corn and French beans, we expect to see increases in the production of these cash crops in the treatment SHGs. The intervention simultaneously relieves marketing and transportation constraints, which should correspond to supply-side cost decreases for farmers in the treatment SHGs. Each of these intermediate steps along the intervention's causal chain is designed to increase the household income.

Shown side-by-side in our table 3, which reproduces AGK's table 4, the replication study results generally follow the original findings. Households in the treatment groups remain more likely to export crops, produce more baby corn and spend less on transportation. AGK highlight that treatment households are 19.2 per cent more likely to adopt export crops in the follow-up period. Our 24.6 per cent result for the same outcome of interest is very similar and of an equal statistical significance level as the original finding.

Table 3: Impacts of DrumNet: ITT OLS, AGK table 4 partial reproduction

	Export Crop	Export Crop AGK	Proportion of Land Devoted to Cash Crops	Proportion of Land Devoted to Cash Crops AGK	Production of Baby Corn (kg)	Production of Baby Corn (kg) AGK	Total Spent in Marketing	Total Spent in Marketing AGK	Logarithm of Household Income	Logarithm of Household Income AGK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: intent-to-treat estimates, OLS										
Post	-0.005	-0.004	-0.096	-0.078	21.455	11,133	0.801	3,567	-0.253	-0.107
	[0.053]	[0.059]	[0.014]***	[0.019]***	[19.496]	[34.775]	[0.791]	[2.133]	[0.099]**	[0.097]
Post x	0.246	0.192	0.061	0.043	86.465	396,735	-1.825	-3,528	0.176	0.089
Treatment	[0.067]***	[0.067]***	[0.017]***	[0.024]*	[41.947]**	[99.607]***	[0.943]*	[1.781]*	[0.113]	[0.110]
Number of observations	1983	1983	1847	1779	1983	1981	1674	1653	1581	1566
R^2	0.203	0.27	0.157	0.13	0.022	0.07	0.026	0.02	0.158	0.16
Mean dependent variable	0.563	0.526	0.515	0.568	43.872	144.6	1.053	1.4	3.498	3.495

Note: Columns 1, 3, 5, 7 and 9 contain replication results. Columns 2, 4, 6, 8 and 10 contain AGK's original results.

In a few important places our results strengthen the evidence for the effectiveness of the intervention. In particular, our results show treatment households being significantly more likely to devote a greater portion of their land to cash crops in the follow-up period.

Similar to AGK, we do not find any significant changes in French bean production, so we present only the baby corn results here. Similar to AGK's results, we find a significant increase in the treated households' likelihood of producing baby corn, albeit with a much smaller coefficient. These results support the theory of change implied by AGK, with the intervention convincing treated households to shift more of their agricultural production into export crops.

Although similar in statistical significance level, our reproduction of baby corn results differ quite substantially in coefficient size from the original results.[14] We find average increases of an extra 86 kilograms in treatment households in the follow-up year. These results are closer to the findings in AGK's working paper (2008), where treatment households average an increase of 396.711 kilograms production of baby corn. Throughout the rest of the paper we correct for misplaced commas in AGK's *production of baby corn*, *production of French beans* and *total spent in marketing* results.

A few notable differences exist between the original findings and the replication results in some of the secondary outcomes in our table 3. While our loan and deposit coefficients are all similar, the statistical significance levels are slightly different. We do not see a significant increase in the likely use of inputs in the follow-up period.[15] But overall, outside of the difference in coefficient size for a few of the outcomes of interest, the original results are generally robust to our replication study.

2.5 Exploring heterogeneous impacts

The original publication explores how the DrumNet package of interventions impacts previous crop exporters versus new adopters. AGK determine that the intervention mainly benefits producers who had not previously grown the export-oriented crops recommended in the intervention. We focus on these results, due to the potentially large policy implications of their finding for future development programmes.

Table 4A: Impacts of DrumNet (prior exporters versus new adopters): ITT OLS, AGK table 6 partial reproduction

	Proportion of Land Devoted to Cash Crops		Proportion of Land Devoted to Cash Crops AGK		Production of Baby Corn (kg)		Production of Baby Corn (kg) AGK	
Grows Export Crops at Baseline	Yes (1)	No (2)	Yes (3)	No (4)	Yes (5)	No (6)	Yes (7)	No (8)
Post	-0.117	-0.059	-0.102	-0.052	13.235	40.277	-18.175	64.590
	[0.020]***	[0.022]**	[0.017]***	[0.034]	[19.918]	[23.447]*	[31.051]	[48.654]
Post x Treat	-0.016	0.104	-0.019	0.086	150.759	30.338	489.112	338.607
	[0.029]	[0.029]***	[0.031]	[0.041]***	[84.862]*	[33.668]	[128.097]**	[104.410]**
Number of observations	895	940	818	909	957	1014	894	1027
R^2	0.19	0.15	0.19	0.14	0.03	0.05	0.10	0.08
Mean dependent variable	0.601	0.432	0.653	0.496	58.274	29.971	147.642	156.560

Note: Columns 1, 2, 5 and 6 contain replication results. Columns 3, 4, 7 and 8 contain original results. Based on our earlier discussion and AGK (2008) we revert some of the publication coefficients to their working paper levels in columns 7 and 8.

[14] Based on Ashraf, Giné and Karlan (2008), we believe the original publication contains typos, as many coefficients are of much smaller magnitudes in an earlier version of the paper. For example, AGK report average increases in baby corn production of 396,735 kilograms in treatment group households. These 396 metric ton increases in baby corn production for treatment communities are difficult to reconcile with the mean baby corn production of 144 kilograms per household reported by AGK in their publication.

[15] As these findings are not central to the theory of change argument we explore, we relegate them to appendix table 6.

The heterogeneous impact findings in our table 4A and table 4B mostly track with the original results. We continue to find, on average, statistically significant increases to the proportion of land treatment households devote to cash crops for new export crop adopters. Our results indicate generally statistically significant changes in regards to previous exporters increasing their baby corn production, but that statistical significance does not carry over to the new export crop adopters in our results.

Table 4B: Impacts of DrumNet (prior exporters versus new adopters): ITT OLS, AGK table 6 partial reproduction

	Total Spent in Marketing (in Ksh1,000)		Total Spent in Marketing (in Ksh1,000) AGK		Logarithm of Household Income		Logarithm of Household Income AGK	
Grows Export Crops at Baseline	Yes (9)	No (10)	Yes (11)	No (12)	Yes (13)	No (14)	Yes (15)	No (16)
Post	0.698	1.512	4.974	2.535	-0.143	-0.364	-0.127	-0.132
	[0.966]	[2.112]	[3.344]	[2.153]	[0.093]	[0.223]	[0.094]	[0.176]
Post x Treat	-2.037	-2.270	-6.488	-1.494	-0.077	0.494	-0.028	0.319
	[1.132]*	[2.481]	[3.319]*	[1.913]	[0.133]	[0.234]**	[0.119]	[0.182]*
Number of observations	883	779	800	793	842	727	764	744
R^2	0.03	0.08	0.03	0.10	0.16	0.20	0.20	0.19
Mean dependent variable	1.085	1.006	1.979	0.768	3.603	3.374	3.641	3.354

Note: Columns 9, 10, 13 and 14 contain replication results. Columns 11, 12, 15 and 16 contain original results. Based on our earlier discussion and AGK (2008) we revert some of the publication coefficients to their working paper levels in columns 11 and 12.

The possible presence of endogenous sorting may bias these heterogeneous impacts, in that households more prone to selling their crops may select themselves into the adoption group. A number of factors may influence crop adoption patterns, including a host of unobservable household and individual attributes. We do not directly address these concerns, although we find little correlational evidence to suggest that richer households sell export crops.

Household income presents the most compelling result in terms of the effectiveness of this intervention. New adopters have an even greater likelihood of significantly increasing their household income in our replication study than in the original paper. These income results reinforce AGK's finding that the intervention caused, on average, a statistically significant increase in the household income of newly commercialised agricultural households.

2.6 Pure replication challenges

Our decision to independently reconstruct the original evaluation from the raw data limits our replication study. These original data do not include a guide to explaining most of the recoding decisions the researchers made during the cleaning process. Researchers clean data for a multitude of reasons, including enumerator errors, data entry mistakes and implementation problems. The code in our evaluation corrects for obvious data outliers, but we keep almost all of the data in their raw form.[16] As seen in our reproduced summary statistics, although the number of observations in the surveys remains very similar, in some instances the magnitudes of the variables of interest vary quite substantially with the published results.

[16] We understand the importance of cleaning data. Eliminating known errors from raw data is an essential step to generating meaningful analysis. Outliers and survey issues arise in research; documenting their identification and elimination would ease future reproduction process.

2.6.1 Data documentation limitations

We encounter difficulties identifying households within and across the survey rounds. We find the unique member identification numbers not entirely consistent throughout the datasets, forcing some data recoding and raising some duplication concerns. The multi-dataset format requires extensive merging throughout the analysis process, making accurate household identification increasingly important.

We also make a number of coding assumptions in our replication process. For example, we assume the extra 500 survey instrument documentation reverses the crop production results, that many of the extra 500 observations are missing decimal points (and thus need to be divided by a factor of 100) and that a number of the variables need to be imputed. We find imputing of missing variables particularly sensitive to the variables we include and the method we choose. We also discover some of the same data to be entered multiple times by different people. Resolving these duplicate data problems is not a straightforward process for us.

We find it easier to reproduce some variables than others. Variables prove difficult to reproduce for a number of reasons. Without knowing the date of the interview, we approximate the number of *months as member in SHG*. The baseline survey does not capture *member of SHG is an officer*, so we obtain this information from follow-up data. Reproducing *value of harvest produced* requires a number of assumptions, from measurement conversions to generating price per unit values. *Total spent in marketing* is the hardest variable for us to reproduce, as it requires household transport costs and assumptions of the number of typical transactions for each household.[17]

2.7 Conversion factors

When designing a survey instrument, researchers must decide how to approach units of measure. Requiring interviewees to report production in standard units makes data analysis easier but may introduce biases into the results if it forces respondents to approximate their production into an unfamiliar unit of measurement. But allowing enumerators to record non-standard units forces researchers to later convert these unique measurement types. These conversions allow for better approximations of production but require accurate agricultural conversion tables.[18]

AGK's survey instruments include local measurement units, which makes conversion of agricultural production and planting responses more difficult to calculate. For example, conversion factors for agricultural production reported in '*gorogoro'* does not exist in standard measurement manuals (ERS 1992) and is apparently a local unit of measure. Kenyan measurement resources, from websites to publications, prove equally unhelpful with converting these measurements into a standard unit of measure.

[17] See appendix table 1 for a full explanation of how we generate the variables. Imputation ultimately played a role in recreating the sample sizes, which were partially guided by the original paper.
[18] For an example of the difficulty around conversion factors, see the World Bank researchers' note on determining how best to convert non-standard measurements in a Malawian household survey (World Bank, undated). And an alternative approach proposed by Verduzco-Gallo, Ecker and Pauw (2014).

Unit measurement conversion factors can influence results. We convert most production and planting area to a standard unit of measurement using AGK's agricultural production conversion tables.[19] As table 5 notes, these conversions enable us to capture almost a quarter of agricultural production. Even with this conversion assistance, we are unable to convert production measured in 'stems'. By not accounting for planting measurements recorded in stems, we could not capture around 9 per cent of the recorded responses.

Table 5: Respondents reporting in standard units of measure

	Baseline period					
	Land planted			Production		
	Standard	Total	Per cent	Standard	Total	Per cent
Original	3691	4002	92.229%	2649	3439	77.028%
Extra 500	1390	1610	86.335%	1294	1598	80.976%
Overall	5081	5612	90.538%	3943	5037	78.281%
	Follow-up period					
	Land planted			Production		
	Standard	Total	Per cent	Standard	Total	Per cent
Original	2567	2837	90.483%	2011	2675	75.178%
Extra 500	1923	2019	95.245%	1626	1993	81.586%
Overall	4490	4856	92.463%	3637	4668	77.913%

Note: Non-standard units of measure for agricultural production include: crates, numbers, bunches, handfuls and other. Non-standard units of measure for amount of land devoted to planting agricultural crops is reported in stems. These calculations exclude 283 missing units of measure for production in the extra 500 baseline. They also exclude 27 production observations recorded with a unit of '0'.

The use of non-standard units of measure for agricultural reporting in the original study touches on a much broader debate about these measurements in the literature (Diskin 1999, Fermont and Benson 2011). Although conversion factors contribute to the difficulties we encounter with reproducing the original evaluation results, we ultimately manage to capture most of the original observations in our analysis.

2.8 Pure replication conclusions

Our overall results are similar, but not the same as those in the original publication. We undertook this replication study understanding the improbability of reproducing the original results exactly. Researchers make numerous decisions during the course of an evaluation, and it is nearly impossible to document each of those decisions. Although our results differ from the original paper in some aspects, we consider the general robustness of our pure replication supportive of continued interest in future agricultural commercialisation projects.

3. Measurement and estimation analysis: power

Our measurement and estimation analysis follows our replication plan in examining the evaluation findings from power and data perspectives. First, we separate the extra 500 households from the original survey participants to check the within- and between-sample balance for both datasets. We then run post-intervention power calculations to lay out observational limitations to the findings, with and without survey data pooling.

[19] AGK made these tables available to us upon request. They noted that they created the tables in conjunction with the Tegemeo Institute at Egerton University in Nairobi.

The publication results rely on pooling the baseline and the extra 500 surveys to measure the impact of the intervention. A number of possible complications accompany this procedure. Enumerators interviewed respondents in the two datasets at different times about their baseline cropland allocations, use of inputs and specific crop production. They collected the original baseline data before the intervention began, when DrumNet had not yet begun operating in the treatment communities and when crops had recently been harvested. Enumerators collected the extra 500 baseline information at the time of follow-up survey. These extra 500 data may include courtesy bias, as the intervention had already begun, or recall bias, as enumerators asked respondents to remember information from a year before.

Table 6: Updated baseline DrumNet summary statistics by dataset: replication results

	Original Data Means			p-Value on t-Test of Difference (2) and (3)	Extra 500 Data Means			p-Value on t-Test of Difference (6) and (7)	p-Value on t-Test of Difference (1) and (5)
	All	Control	Treatment		All	Control	Treatment		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Member									
Age of member	43.17	41.98	43.83	0.07*	40.71	35.56	42.52	0.00***	0.00***
	(13.22)	(13.20)	(13.20)		(11.23)	(8.77)	(11.45)		
Literacy	0.89	0.88	0.89	0.60	0.95	0.99	0.93	0.03**	0.00***
	(0.32)	(0.33)	(0.31)		(0.22)	(0.10)	(0.25)		
(i) Member of SHG is an officer (1=yes)	0.17	0.16	0.18	0.38	0.08	0.08	0.08	0.94	0.00***
	(0.38)	(0.36)	(0.39)		(0.27)	(0.27)	(0.27)		
Land									
Proportion of land that is irrigated	0.32	0.35	0.30	0.04**	0.45	0.41	0.46	0.14	0.00***
	(0.31)	(0.29)	(0.32)		(0.30)	(0.31)	(0.30)		
Total landholdings	2.28	2.37	2.23	0.33	1.83	1.75	1.86	0.61	0.00***
(Acres)	(1.95)	(2.09)	(1.87)		(1.84)	(1.27)	(2.01)		
Proportion of land devoted to cash crops	0.53	0.55	0.52	0.26	0.52	0.57	0.50	0.04**	0.44
	(0.27)	(0.26)	(0.27)		(0.30)	(0.32)	(0.29)		
Production									
Grows export crop	0.57	0.65	0.53	0.00***	0.30	0.36	0.28	0.13	0.00***
(1=yes)	(0.50)	(0.48)	(0.50)		(0.46)	(0.48)	(0.45)		
Used hired labour	0.29	0.29	0.29	0.98	0.39	0.40	0.38	0.73	0.00***
(1=yes)	(0.45)	(0.45)	(0.45)		(0.49)	(0.49)	(0.49)		
Uses machinery and/or animal force	0.07	0.10	0.05	0.01***	0.00	0.00	0.00	.	0.00***
(1=yes)	(0.26)	(0.31)	(0.23)		(0.00)	(0.00)	(0.00)		
Production of French beans (in 1,000 kg)	1.02	0.66	1.21	0.31	0.20	0.20	0.21	0.94	0.03**
	(6.71)	(1.34)	(8.24)		(1.06)	(0.53)	(1.18)		
Production of baby corn (in kg)	4.84	13.64	0.18	0.14	2.57	2.35	2.65	0.92	0.72
	(114.06)	(193.73)	(3.74)		(23.79)	(21.69)	(24.49)		
Use of inputs	0.98	0.98	0.98	0.63	0.97	0.99	0.96	0.21	0.22
	(0.14)	(0.13)	(0.15)		(0.18)	(0.11)	(0.20)		

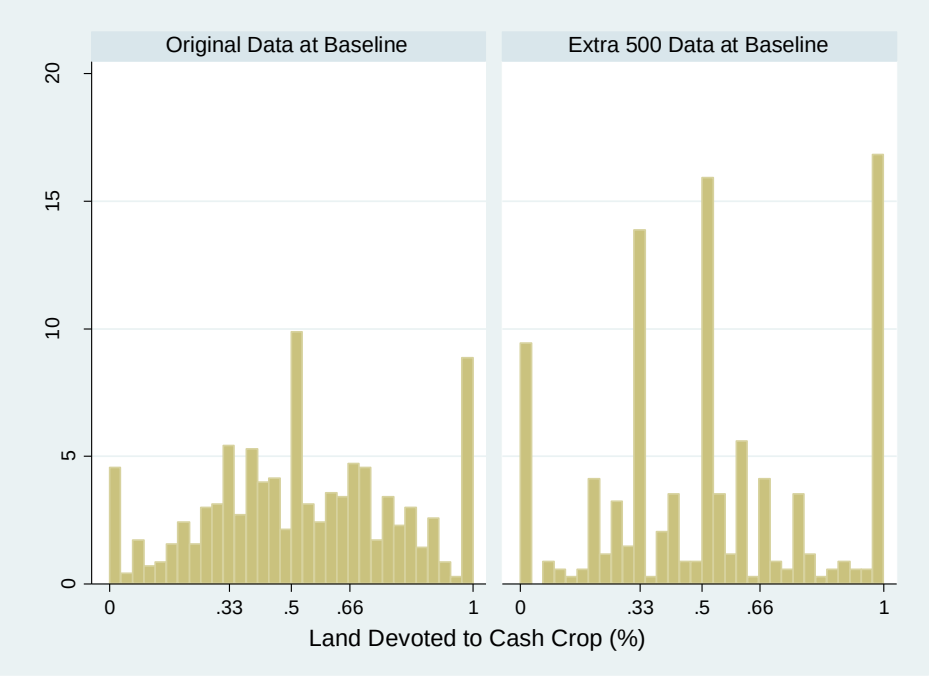
Note: We break results out by data source, with all results coming from the replication study. Column 9 compares all of the 'original' data to all of the extra 500 data among select variables at the baseline time period.

Given the differences in the baseline data collection timelines, we examine the datasets separately. Table 6 shows the balance between the original baseline data and the extra 500 recall data. When comparing table 6 with the balance shown in table 2A, similarities and differences emerge. *Age* continues to be statistically significantly imbalanced across all of our datasets. The other variables highlighted in table 2A, *export crops* and *uses machinery/animal force*, are only significantly imbalanced statistically in the original dataset. *Literacy* and *proportion of land devoted to cash crops* are only significantly imbalanced in the extra 500 data.

We identify a number of statistically significant imbalances when comparing the balance between the two datasets. Following Bruhn and McKenzie (2009), we combine the data and conduct an F-test on the joint orthogonality of the variables in relation to treatment. With an F value of 4.13, the test finds the collective contribution of the variables in the balance tables to be significantly different from each other at the 1 per cent level when comparing the control and the treatment households.

Of all of the evaluation's outcome variables, we focus our analysis on the one arguably hardest to precisely recall, *proportion of land devoted to cash crops*. To better understand differences between the two surveys, we separate the responses by the dataset and treatment status in the histogram plots in figure 1. As the recall period refers to the baseline period, we only present the baseline data results.

Figure 1: Baseline data distribution of land devoted to cash crops, by dataset



The distribution of the baseline responses for *proportion of land devoted to cash crops* differ quite substantially between the two datasets. We find a wide distribution of answers in the original dataset. Less variation exists in responses to the *proportion of land devoted to cash crops* question in the extra 500 data. The clumpier distribution of the extra 500 data around one third and one half suggests possible recall approximations in these responses. We formally test the equality of the distributions with a Kolmogorov–Smirnov test, finding a statistically significant difference between the two groups at a 1 per cent level (Stephens 1992).

Because of our concerns with the recall data, we reproduce the main evaluation results excluding the extra 500 data. Although we continue to find similar coefficient size between the combined and original samples, some of the key results lack statistical significance in the original-only sample. The extra 500 data appear to at least drive some of the main results, especially in the *proportion of land devoted to cash crop* variable coefficient, which switches signs after excluding the extra 500 data.

Table 7: Impacts of DrumNet: ITT OLS by dataset

	Export crop, original only	Export crop, entire sample	Land devoted to cash crops, original only	Land devoted to cash crops, entire sample	Production of baby corn (kg), original only	Production of baby corn (kg), entire sample	Total spent in marketing, original, original only	Total spent in marketing, entire sample	Log of household income, original only	Log of household income, entire sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: intent-to-treat estimates, OLS										
Post	-0.089	-0.005	-0.144	-0.096	21.894	21.455	1.535	0.801	-0.222	-0.253
	[0.082]	[0.053]	[0.020]***	[0.014]***	[24.784]	[19.496]	[1.201]	[0.791]	[0.100]**	[0.099]**
Post x	0.227	0.246	-0.006	0.061	114.220	86.465	-2.518	-1.825	0.099	0.176
Treatment	[0.106]**	[0.067]***	[0.027]	[0.017]***	[70.791]	[41.947]**	[1.376]*	[0.943]*	[0.125]	[0.113]
Number of Observations	1252	1983	1117	1847	1252	1983	1279	1674	1186	1581
R^2	0.290	0.203	0.244	0.157	0.036	0.022	0.030	0.026	0.171	0.158
Mean dependent variable	0.613	0.563	0.497	0.515	50.887	43.872	1.252	1.053	3.497	3.498

Note: We present the main outcomes of interest from the main ITT analysis excluding the extra 500 data in columns 1, 3, 5, 7 and 9. For comparison purposes, we also reproduce the entire sample results from our table 3 above in columns 2, 4, 6, 8 and 10.

Sample size constraints restrict our ability to make definite statements about the results from the separate datasets. AGK specifically collected the extra 500 data to offset their power concerns with the original size of the sample. To better understand these constraints, we determine power calculations for the evaluation.

We approach the power calculations from the baseline sample to determine the necessary size of the sample and to possibly see a statistically significant change in the outcomes of interest. As we run post-intervention power calculations, we calculate the actual intra-cluster correlation coefficient (ICC) for each outcome of interest, the standard deviation of the outcome variable, and the proportion of the outcome variable explained by the covariates. Our post-evaluation ICCs provide researchers with guidance for future evaluations designed around these outcomes of interest.[20]

Initially, we examine the entire sample to the minimum detectable effect sizes to detect behaviour changes. Based on the traditional social science requirements of 80 per cent power and a statistical significance level of 5 per cent, most of the coefficients exceed the minimal detectable effect level in table 8. Due to our recall bias concern, we then exclude the extra 500 data from the analysis in the same table.

We focus our interest on the *log of household income* variables in the power tables due to the centrality of this outcome to the intervention's theory of change and its statistical insignificance in both the original publication and the replication study. As *log of household income* ultimately measures the effectiveness of the intervention, we give it extra attention.

Our power calculation results show, unsurprisingly given the results, that the *log of household income* variable does not contain enough observations to detect statistical significance. Given the entire sample size and the adoption rate seen in the study, table 8 shows there would have had to be a minimum change of 0.329 in this variable to be able to detect a statistically significance change between the control and treatment groups at the 5 per cent level with 80 per cent confidence.

[20] Our purpose in running post-evaluation power calculations is to determine required sample sizes. See Hoenig and Heisey (2001) and Lenth (2007) for a wider discussion of issues with post-hoc power analysis. See Wood and Djimeu (2014) for more information on the importance of documenting power calculations.

Table 8: Post-intervention power calculations, baseline by dataset

Outcome of interest	Including extra 500 dataset	Intra-cluster correlation coefficient	Minimum detectable effect size	Change in outcome of interest, our replication study	Change in outcome of interest, AGK publication
	(1)	(2)	(3)	(4)	(5)
Proportion of land devoted to cash crops	Yes	0.101	0.099	0.061	0.043
	No	0.105	0.105	-0.006	n/a
Baby corn production	Yes	0.013	20.441	86.465	396.735
	No	0.018	28.192	114.220	n/a
French bean production	Yes	0.047	1.488	2.232	1.611
	No	0.037	1.892	3.119	n/a
Total spent in marketing	No	0.024	1.545	-1.825	-3.528
Log of household income	No	0.050	0.329	0.176	0.089

Note: We use Djimeu and Houndolo (2015) to conduct these power calculations, assuming a standard statistical significance level of 5 per cent and a .80 power of the test. We calculate the post-intervention ICCs in Stata, separating the results by inclusion of either baseline samples or only the original data set. This separation does not apply to *time spent in marketing* and *log of household income*, as baseline information was not collected for these variables. AGK did not publish results without the extra 500 data, so we report those results as not applicable in the final column. Based on our earlier coefficient size discussion and Ashraf *et al.* (2008) we revert some of the publication coefficients to their working paper levels.

After calculating the effect size, still focusing on *log of household income,* we determine the required size of future evaluations to be able to test for statistical significance. We perform a back-of-the-envelope calculation assuming the same number of SHGs, the same average number of interviewees per SHG and that our replication results are the true change in household income. Under these assumptions we estimate that future evaluations would need to include approximately 2,500 people to be able to detect a statistically significant increase in the *log of household income* among treatment households.

4. Theory of change analysis: reanalysis and alternative heterogeneity analysis

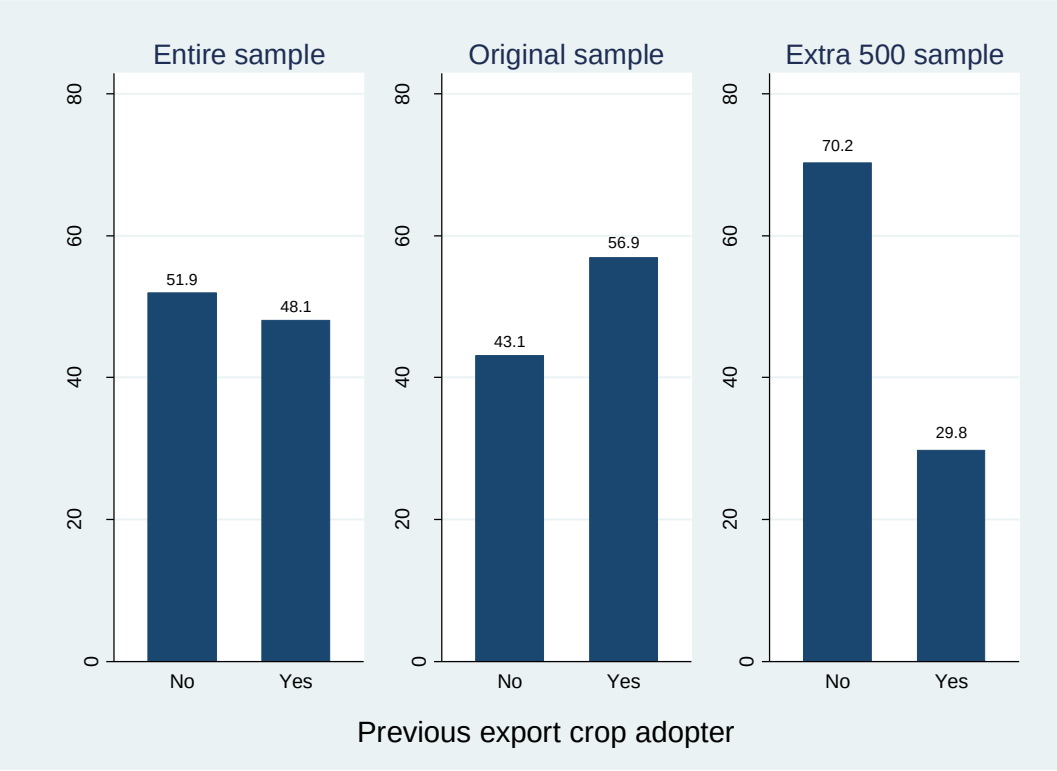
The original findings indicate that the intervention influences new farmers adopting export crops differently than farmers who were growing these crops for export before the intervention began, suggesting that farmers react differently to the package of interventions based on their existing agricultural practices. AGK's findings could lead policymakers to target only subsistence farmers for agricultural commercialisation projects. Our theory of change analysis examines AGK's original heterogeneous impact results and explores two alternative approaches to defining previous adopters that may improve future interventions targeting efforts.

AGK's heterogeneous impact results may lead to quite striking policy implications, notably that policymakers should focus their efforts on subsistence farmers not already growing export crops. We explore the hypothesis that the DrumNet intervention encourages formerly isolated farmers to work within established markets. Under this theory, the specific export crop encouragement is less important than the commercialisation of smallholder farmers. This alternative hypothesis would alter the policy recommendations somewhat, away from encouraging specific export crop production and towards supporting farmers to engage with markets.

In our opinion, DrumNet's success with subsistence farmers may stem from the intervention's focus on factors that had previously prevented these farmers from growing more valuable crops. A large body of literature finds subsistence crop production a typically inefficient use of land, with labour-intensive commercial crops potentially allowing for significant welfare gains through comparative advantage in lower wage costs.[21] Relating to market failures, subsistence farmers may not adopt export crops due to a lack of established relationships with traders. In addition, these same farmers may fear the unknown effects of adoption, possibly because of a need for knowledge of advanced agricultural practices or uncertainty around increased household dependence on outside markets for food purchases.

AGK do not explicitly state their theory of change or explain why they choose to focus on possible 'previous adopter' heterogeneous impacts of the intervention. As seen in figure 2, large differences exist between the per cent of previous adopters of export crops in the original baseline data and the extra 500 baseline recall data. We find 57 per cent of farmers having adopted export crops in the original baseline, compared with only 30 per cent of the extra 500 sample recalling being of similar status. The marked difference in the baseline status of the two samples questions the interpretability of the pooled heterogeneous impacts.

Figure 2: Previous adopters of export crops, by dataset



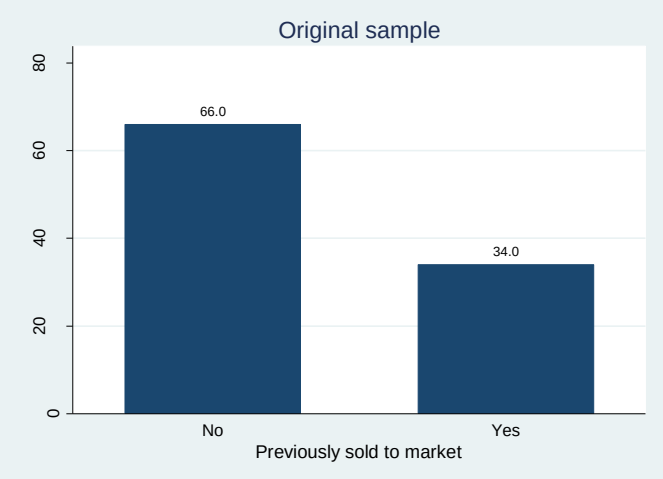
In addition to our concerns over differences between the two datasets, we also explore the robustness of the heterogeneous outcomes findings in relation to alternative definitions of previous adopters. AGK define previous adopters rather narrowly as those farmers who previously grew the three DrumNet target crops. We expand AGK's definition in two directions.

[21] See Wood *et al.* (2014) for a literature review around subsistence versus cash crop production.

We first test the robustness of AGK's heterogeneous impact findings by expanding the export crop definition to include farmers who grew any cash crops before the intervention. AGK's definition of cash crops includes various fruits, vegetables and nuts, along with other export crops such as tea and coffee. Under this alternative previous adopter definition, we classify anyone who grew any of these crops before the baseline survey as a previous adopter. Unfortunately, this alternative definition captures almost all of the households in the samples.

We also explore the alternative hypothesis, from our replication plan, that farmers who previously sold to markets might better capture heterogeneous impacts of the intervention. Our *previously sold to market* binary variable takes a value of 1 for households that had previously sold crops to the village or distant market.[22] We note in figure 3 that 66 per cent of households in the original sample had not previously sold to market at the baseline period. As the extra 500 sample does not include this recall information for this baseline variable, we use observations only from the original dataset for this alternative definition.

Figure 3: Previously sold to market, original dataset



Rerunning the original heterogeneous impact estimations with our alternative previous adopter variables creates different sets of results. First, in panel A of tables 9A and 9B, we separate the original heterogeneous impact results into the full sample and the sub-sample that only includes the original baseline households. Then we replace the before-intervention cash crop adopter category with our *cash crop* and our *sold crops at a market* variables in panels B and C of these tables.[23]

Focusing on panel A in the two tables, we find that limiting the sample to only the households included in the original baseline survey generally reduces the magnitude of the coefficients on the outcomes of interest. These results would support a hypothesis that the extra 500 data represent a different subsample of households than the original baseline households. The findings would also support the possibility of a courtesy bias in the extra

[22] We do not include crop sales to traders or to auctions in our *sold to market* variable. A number of these sales are classified as 'other', which we also consider outside of market sales.
[23] We also planned to explore distance to market differences, but privacy concerns prevented AGK from sharing household geographic location data about the sample populations.

500 household responses (Crawford 1997). In the courtesy bias scenario, treatment households, having been exposed to the intervention before being asked recall baseline questions, might tailor their answers to demonstrate the effectiveness of the intervention.[24]

Table 9A: Impacts of DrumNet, ITT OLS heterogeneous outcomes, by baseline characteristics and dataset

	Proportion of Land Devoted to Cash Crops, entire sample		Proportion of Land Devoted to Cash Crops, excluding extra 500		Production of Baby Corn (kg), entire sample		Production of Baby Corn (kg), excluding extra 500	
	Yes	No	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Binary baseline heterogeneous variable, Previously grew export crop								
Post	-0.117	-0.059	-0.136	-0.140	13.275	40.562	7.941	64.338
	[0.020]***	[0.022]**	[0.026]***	[0.038]***	[19.923]	[23.491]*	[19.921]	[38.429]
Post x Treat	-0.016	0.104	-0.053	0.042	151.484	30.233	219.141	-8.010
	[0.029]	[0.029]***	[0.038]	[0.044]	[84.854]*	[33.686]	[126.742]	[37.074]
Number of observations	895	940	646	471	957	1014	707	545
R^2	0.19	0.15	0.22	0.2	0.03	0.05	0.06	0.05
Mean dependent variable	0.601	0.432	0.582	0.381	58.473	30.070	1.956	3.598
Panel B: Binary baseline heterogeneous variable, Previous cash crop producer								
Post	-0.098	0.253	-0.157	0.262	21.936	8.935	23.451	29.001
	[0.019]***	[0.069]***	[0.023]***	[0.094]**	[18.369]	[12.021]	[26.067]	[34.026]
Post x Treat	0.047	0.044	0.001	-0.207	91.139	53.902	117.249	35.683
	[0.023]**	[0.090]	[0.030]	[0.107]*	[44.973]*	[27.995]*	[73.712]	[34.307]
Number of observations	1665	170	1053	64	1786	185	1174	78
R^2	0.18	0.64	0.23	0.86	0.03	0.39	0.03	0.68
Mean dependent variable	0.542	0.242	0.514	0.222	2.210	0.449	2.789	0.893
Panel C: Binary baseline heterogeneous variable, Previously sold to market								
Post	n/a	n/a	-0.122	-0.158	n/a	n/a	5.789	32.479
			[0.029]***	[0.025]***			[25.597]	[35.312]
Post x Treat			-0.008	-0.001			63.329	131.449
			[0.055]	[0.033]			[31.948]*	[100.596
Number of observations			390	390			390	390
R^2			0.29	0.28			0.18	0.04
Mean dependent variable			0.501	0.495			34.630	59.573

Note: All columns are replication results. Export crop refers to households who sold French beans and baby corn before the intervention began. Cash crop refers to households who sold any crops defined as cash crops, outside of those typically grown by subsistence farmers, before the intervention began. Sold to market refers to households who report selling crops to a village or distant markets in the baseline period. Extra 500 households were not asked to recall their baseline markets sales, so we list those results as not applicable in the table.

When examining heterogeneous impacts from a previous cash crop producer perspective, our results somewhat follow the export crop producer findings. We discount all of these results, reported in Panel B of tables 9A and 9B, because the vast majority of households in the samples previously grew cash crops. Thus, while cash crop production theoretically fulfills the requirements for an alternative means to measure heterogeneous impacts of the intervention on subsamples of the population, the sample does not include enough non-cash crop producing households to enable this analysis to produce accurate results in practice.

[24] As we discuss in our measurement and estimation analysis, the reduced sample sizes from excluding the extra 500 data will affect our ability to detect statistical significance in all results in columns 3, 4, 7 and 8 in tables 9A and B.

Table 9B: Impacts of DrumNet, ITT OLS heterogeneous outcomes, by baseline characteristics and dataset

	Total Spent in Marketing (in Ksh 1,000), entire sample		Total Spent in Marketing (in Ksh 1,000), excluding extra 500		Logarithm of Household Income, entire sample		Logarithm of Household Income, excluding extra 500	
	Yes	No	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Binary baseline heterogeneous variable, Previously grew export crop								
Post	0.698	1.512	1.306	2.220	-0.143	-0.364	-0.117	-0.379
	[0.966]	[2.112]	[1.200]	[3.089]	[0.093]	[0.223]	[0.101]	[0.211]*
Post x Treat	-2.037	-2.270	-2.775	-2.837	-0.077	0.494	-0.113	0.427
	[1.132]*	[2.481]	[1.421]*	[3.528]	[0.133]	[0.234]**	[0.141]	[0.237]*
Number of observations	883	779	734	545	842	727	693	493
R^2	0.03	0.08	0.04	0.09	0.16	0.2	0.17	0.24
Mean dependent variable	1.085	1.006	1.233	1.279	3.603	3.374	3.622	3.320
Panel B: Binary baseline heterogeneous variable, Previous cash crop producer								
Post	0.999	-0.142	1.807	-0.289	-0.255	-0.113	-0.243	0.117
	[0.865]	[0.879]	[1.323]	[0.950]	[0.109]**	[0.258]	[0.122]*	[0.343]
Post x Treat	-2.124	0.470	-2.855	0.464	0.192	0.133	0.132	-0.327
	[1.030]**	[1.051]	[1.508]*	[1.215]	[0.125]	[0.366]	[0.143]	[0.456]
Number of observations	1479	183	1174	105	1393	176	1088	98
R^2	0.03	0.16	0.03	0.18	0.16	0.34	0.16	0.42
Mean dependent variable	1.141	0.294	1.328	0.412	3.533	3.211	3.520	3.238
Panel C: Binary baseline heterogeneous variable, Previously sold to market								
Post	n/a	n/a	1.462	1.564	n/a	n/a	-0.001	-0.325
			[2.811]	[1.499]			[0.174]	[0.167]*
Post x Treat			-3.671	-1.954			-0.377	0.321
			[2.805]	[1.758]			[0.210]*	[0.187]*
Number of observations			390	390			390	390
R^2			0.05	0.08			0.25	0.19
Mean dependent variable			2.011	0.859			3.553	3.467

Note: All columns are replication results. Export crop refers to households who sold French beans and baby corn before the intervention began. Cash crop refers to households who sold any crops defined as cash crops, outside of those typically grown by subsistence farmers, before the intervention began. Sold to market refers to households who report selling crops to a village or distant markets in the baseline period. Extra 500 households were not asked to recall their baseline markets sales, so we list those results as not applicable in the table.

Finally, when we change the binary *previous adopter* baseline variable to *previously sold to market*, we find very similar results with the original heterogeneous impact analysis. As we note above, the extra 500 households were not asked to recall this question, so we cannot present entire sample heterogeneous impact results.

Specifically, we see average increases in baby corn production by treatment households, and marketing costs tend to reduce. Of particular interest, we continue to find dichotomous *log of household income* results. Following this alternative approach to measuring heterogeneous impacts, treatment households that previously sold to markets average statistically significant decreases in the income measurement, whereas we find the opposite result for treatment households who had previously not sold to a market.

The relationship between different channels of entering the agricultural production value chain remains an area for future research. We believe previously selling crops to market to be more relevant than producing specific export crops, as markets should dictate the highest

prices for the crops with the most demand. But incomplete markets may interrupt this natural process. Fan *et al.*'s (2013) recent work on helping smallholder farmers transition from subsistence farming to more profitable agricultural strategies provides policy options to encourage smallholder participation in the commercial food chain. These *sold to market* and *previous adopter* results suggest avenues for additional agricultural commercialisation research, to support better targeted future agricultural commercialisation interventions.

5. Limitations

We see numerous limitations to our replication study, a number of which we have outlined above. Some additional limitations are facts of the evaluation, whereas others might be addressed in future replication work. Overall, we do not consider any of these limitations overly detrimental to our study results.

We encounter a number of difficulties with reproducing the original evaluation. Our incomplete knowledge of the survey collection procedures limits our replication study. We independently reconstruct the evaluation from the raw data instead of use the existing .do files, which creates many research decisions that may influence the study results. Identifying households through the different rounds of the survey and identifying outliers within the data prove particularly hard for us throughout the replication process.

Time and data limitations prevent us from addressing all of the anticipated activities in our replication plan. We do not examine potential contamination concerns in our study or possible gender components of the research. We also hoped to explore the surprising positive coefficient on the credit treatment group for *total spent in marketing* in AGK's table 5. As the intervention included marketing assistance, we would expect the marketing expenditures coefficient to have a negative sign. But our inability to accurately disentangle treatment assignment discourages us from delving further into the heterogeneous treatment impact tables. We do report a negative coefficient for total spent in marketing in our reproduction of this finding in our appendix table 7. AGK (2008) also find a negative coefficient on *total spent in marketing* in their table 3, so the published result may simply be a typo.

6. Conclusions

Our replication study generally supports the original publication results and, in turn, the concept of agricultural commercialisation. Our results find evidence in favour of targeting households not yet participating in agricultural markets, although this result would benefit from additional research. Our findings suggest that the intervention may benefit households by encouraging them to participate in market-based economies. We consider the original study and the replication results to support further research into the viability of increasing agricultural commercialisation among smallholder farmers in the developing world.

We find our data pooling and power analysis results most striking. The differences between the two baselines make us question the interpretability of the study's findings because of its reliance on pooling these samples. In terms of the statistically insignificant increases to household income, our analysis suggests future evaluation would need to substantially increase sample sizes, on the order of quadrupling the original sample size, to be able to detect a statistically significant difference between treatment and control groups in this regard. We encourage future researchers to provide greater insights regarding the effectiveness of the intervention.

AGK encounter numerous constraints with this evaluation, from small initial sample sizes to unanticipated trade barriers. These internal and external issues limit the scope of the findings. Most notably, the short time frame of this evaluation probably undervalues this package of agricultural commercialisation interventions. In future studies in which the theory of change is so complex, we strongly recommend researchers conduct evaluations with larger sample sizes to ensure adequate power for the main outcomes of interest and any additional heterogeneous impact analysis.

Our replication study provides a number of avenues for further agricultural commercialisation research. One particularly promising research area is disentangling the relationship between different channels of entering the agricultural production value chain. Another connected area for future research is the potential for targeting agricultural commercialisation interventions. We consider the necessity of working with smallholder groups or exclusively with subsistence farmers within agricultural commercialisation interventions to be open research questions. We hope this replication study will encourage future work in this field.

Appendix A: Tables

Appendix Table 1: Variable definition and construction

Age of member	Age of the SHG member. Respondents reported their ages in the baseline survey. Because the follow-up survey was conducted one year after the baseline, we increase the age by one for the age in the follow-up. To recover 67 observations we replaced missing age variables with the average age of the sample.
Literacy	Self-reported ability to read and write. We assume that the literacy of the respondent does not change over the year and use the literacy reported in the baseline for the follow-up data.
Risk tolerance	It was not evident from the survey instrument or the publication how to calculate risk tolerance. Through communications with the original authors we used their method to calculate the risk tolerance based on the hypothetical questions (which are not included in the survey instrument). We assume that risk tolerance is unchanged throughout the survey periods.
Months as member in SHG	Number of months since the respondent became a SHG member. Since the follow-up survey was conducted 13 months after the baseline survey, we add 13 months to the number of months reported at the baseline for follow-up data.
Member of SHG is an officer	Dummy variable with value 1 if respondent was an officer of the SHG. This variable was only asked in the follow-up survey and proved difficult to create for a number of households. We assumed missing observations were not officers in our analysis.
Deposit in a formal bank	Dummy variable with value 1 if any member of the respondent's household has deposits in a formal bank.
Loan from formal institutions	Dummy variable with value 1 if any member of the respondent's household obtained credit from a formal institution such as AFC, commercial, coffee co-op (SACCO) or KTDA.
Total household income	Sum of wages from agricultural labour, wages or salaries from other work, non-farm self-employment, sale of crops, sale of livestock, poultry and dairy, remittances from family members, pension, gifts or social assistance and other income.
Total agricultural income	Sum of wages from agricultural labour, sale of crops and sale of livestock, poultry and dairy.
Uses hired labour	Dummy variable with a value of 1 if the household used hired labour during the last season.
Grows export crops	Dummy variable with a value of 1 if the household grows French beans, baby corn or passion fruit.
Use of inputs	Dummy variable that equals 1 if the household used manure, chemical fertiliser or pesticides for crop production.
Value of harvested produce	Sum of all crops in each plot cultivated of the total amount harvested times the price per unit in a typical transaction in Ksh1,000. The price per unit is calculated by dividing the value of each crop sold in a typical transaction by the amount sold in each transaction. For each household, we average the price per unit by each crop so that we have a unique price per unit for each crop sold by each household. By

	summing the quantity of each crop harvested multiplied by the crop's price per unit, we calculate the value of harvested produce by each household. Since we do not have market data for the extra 500, there are low observation numbers compared with other variables.
Harvest yield per acre	Value of harvest divided by total landholdings (acres) in Ksh100,000. We divide the value of harvested produce by 100 to create value in Ksh100,000 and then divide by the total landholdings.
Proportion of land that is irrigated	Total land that uses some source of irrigation other than rain, divided by the total area of land. We replace the value to 1 for households that report the total area of irrigated land is larger than the total area of land. We also assume missing observations did not irrigate their land. Area of land irrigated and total landholdings appear to be misreported in the follow-up 500 survey. We correct for this by dividing these variables by 100.
Total landholdings (acres)	Total landholdings in acres. We were unable to determine how to convert landholdings reported in 'stem' and thus lost them from our sample. We assume missing observations hold no land. Area of land irrigated and total landholdings appear to be misreported in the follow-up 500 survey. We correct for this by dividing these variables by 100.
Per cent of land devoted to cash crops	We consider cash crops as all non-subsistence crops (beans, maize, potatoes and kale). We calculate the total land devoted to cash crops using the crop types that do not include subsistence crops and divide by the total area of land.
Production of French beans	Sum of the harvested amount of French beans, divided by 1,000 to express the value in 1,000 kg.
Production of baby corn	Sum of the harvested amount of baby corn in kg.
Sells to market	A dummy variable that equals 1 if the respondent reports that s/he sold crops at a village market or a distant market.
Total spent in marketing	Total cost of transport of a typical transaction times the number of transactions that required transportation. Each household reports the total cost of transport by crop type. We multiply the reported cost by the number of sales of each crop and sum the cost at household-level. Since we assume that the transportation cost is zero for those who travelled on foot, we replace the transportation cost to zero for those observations. We also assume households make a maximum of 100 transactions.

Appendix Table 2A: Number of SGHS, replication and original results

	Baseline SHG Number	Follow-up SHG Number	Baseline AGK SHG Number	Follow-up AGK SHG Number
	(1)	(2)	(3)	(4)
Original intervention treatment assignment rule				
Control	12	12	12	12
Credit	12	12	12	12
No Credit	12	11	12	12
Total	36	35	36	36
Updated intervention treatment assignment rule				
Control	12	12	12	12
Credit	13	12	12	12
No Credit	11	11	12	12
Total	36	35	36	36
Follow-up intervention treatment assignment rule				
Control	14	13	12	12
Credit	11	11	12	12
No Credit	11	11	12	12
Total	36	35	36	36

Note: Replication study results are reported in columns 1 and 2. Original results are reported in columns 3 and 4.

Appendix Table 2B: SGHS and Observations by Assignment Rule

	Baseline Number of Observations	Follow-up Number of Observations	Total Number of Observations	Baseline AGK Number of Observations	Follow-up AGK Number of Observations	Total AGK Number of Observations
	(5)	(6)	(7)	(8)	(9)	(10)
Original intervention treatment assignment rule						
Control	361	298	659	367	303	670
Credit	397	344	741	373	316	693
No Credit	352	306	658	377	337	714
Total	1110	948	2058	1117	956	2073
Updated intervention treatment assignment rule						
Control	361	298	659	367	303	670
Credit	397	344	741	373	316	693
No Credit	352	306	658	377	337	714
Total	1110	948	2058	1117	956	2073
Follow-up intervention treatment assignment rule						
Control	381	319	700	367	303	670
Credit	373	321	694	373	316	693
No Credit	356	308	664	377	337	714
Total	1110	948	2058	1117	956	2073

Note: Replication study results are reported in columns 5─7. Original paper observations vary by variable. We report the largest number of observations from each round of the original findings, as found in AGK appendix table 2.

Appendix Table 3: Number of observations using 'updated' assignment, AGK appendix table 2 reproduction

	Baseline					Follow-up				
	All	Control	Treatment	Credit	No Credit	All	Control	Treatment	Credit	No Credit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Member										
Age of member	1110	361	749	397	352	948	298	650	344	306
Literacy	1110	361	749	397	352	948	298	650	344	306
Risk tolerance	726	261	465	217	248	948	298	650	344	306
Months as member in SHG	726	261	465	217	248	948	298	650	344	306
Member of SHG is an officer (1=yes)	1110	361	749	397	352	948	298	650	344	306
Deposit in a formal bank (1=yes)	717	254	463	215	248	948	298	650	344	306
Loan from formal institution (1=yes)	726	261	465	217	248	948	298	650	344	306
Logarithm of total annual household income	707	255	452	214	238	874	290	584	331	253
Number of household members	722	260	462	215	247	948	298	650	344	306
Land										
Harvest yield per acre (in Ksh 100,000)	726	261	465	217	248	920	290	630	329	301
Proportion of land that is irrigated	1110	361	749	397	352	948	298	650	344	306
Total landholdings (Acres)	1110	361	749	397	352	948	298	650	344	306
Proportion of land devoted to cash crops	1037	327	710	375	335	809	274	535	319	216
Production										
Grows export crop (1=yes)	1037	327	710	375	335	945	297	648	343	305
Sells to market (1=yes)	726	261	465	217	248	948	298	650	344	306
Used hired labour (1=yes)	1110	361	749	397	352	948	298	650	344	306
Uses machinery and/or animal force (1=yes)	1110	361	749	397	352	948	298	650	344	306
Value of harvested produce (in Ksh 1,000)	699	242	457	212	245	948	298	650	344	306
Production of French beans (in 1,000 kg)	1037	327	710	375	335	945	297	648	343	305
Production of baby corn (in kg)	1037	327	710	375	335	945	297	648	343	305
Total spent in marketing (in Ksh 1,000)	726	261	465	217	248	948	298	650	344	306
Use of inputs	1037	327	710	375	335	809	274	535	319	216
Follow-up										
Proportions of respondents reached at follow-up	0.85	0.83	0.87	0.87	0.87					

Appendix Table 4: Baseline summary statistics using 'updated' intervention treatment assignments, AGK table 2 reproduction

	Means			p-Value on t-Test of Difference (2) and (3)	Means		p-Value on F-test for (5) and (6)
	All (1)	Control (2)	Treatment (3)	(4)	Credit (5)	No Credit (6)	(7)
Member							
Age of member	42.32	40.20	43.34	0.00***	43.19	43.50	0.00***
	(12.62)	(12.46)	(12.57)		(12.84)	(12.28)	
Literacy	0.91	0.91	0.91	0.91	0.91	0.91	0.99
	(0.29)	(0.29)	(0.29)		(0.29)	(0.29)	
Risk tolerance	0.37	0.37	0.37	0.85	0.36	0.38	0.89
	(0.42)	(0.42)	(0.42)		(0.42)	(0.43)	
Months as member in SHG	53.13	58.33	50.20	0.01***	48.79	51.44	0.02**
	(40.07)	(44.68)	(36.96)		(33.37)	(39.86)	
Member of SHG is an officer (1=yes)	0.14	0.14	0.14	0.75	0.12	0.17	0.18
	(0.35)	(0.34)	(0.35)		(0.33)	(0.37)	
Deposit in a formal bank (1=yes)	0.70	0.73	0.69	0.25	0.72	0.66	0.22
	(0.46)	(0.45)	(0.46)		(0.45)	(0.47)	
Loan from formal institution (1=yes)	0.04	0.06	0.03	0.05**	0.06	0.00	0.00***
	(0.19)	(0.23)	(0.17)		(0.23)	(0.06)	
Logarithm of total annual household income	3.53	3.63	3.47	0.10*	3.71	3.27	0.00***
	(1.17)	(1.16)	(1.18)		(1.15)	(1.17)	
Number of household members	4.59	4.52	4.63	0.50	4.72	4.56	0.57
	(2.09)	(2.10)	(2.08)		(2.24)	(1.94)	
Land							
Harvest yield per acre (in Ksh 100,000)	1.02	0.80	1.15	0.33	1.41	0.92	0.32
	(4.63)	(6.02)	(3.62)		(4.98)	(1.69)	
Proportion of land that is irrigated	0.37	0.37	0.36	0.84	0.40	0.32	0.00***
	(0.31)	(0.29)	(0.32)		(0.32)	(0.32)	
Total landholdings (Acres)	2.13	2.20	2.09	0.36	2.01	2.18	0.33
	(1.93)	(1.91)	(1.93)		(1.94)	(1.93)	
Proportion of land devoted to cash crops	0.53	0.55	0.51	0.03**	0.53	0.50	0.03**
	(0.28)	(0.28)	(0.28)		(0.26)	(0.29)	
Production							
Grows export crop (1=yes)	0.48	0.58	0.44	0.00***	0.49	0.38	0.00***
	(0.50)	(0.49)	(0.50)		(0.50)	(0.49)	
Sells to market (1=yes)	0.34	0.36	0.33	0.53	0.29	0.37	0.23
	(0.47)	(0.48)	(0.47)		(0.46)	(0.48)	
Used hired labour (1=yes)	0.32	0.32	0.32	0.88	0.34	0.30	0.48
	(0.47)	(0.47)	(0.47)		(0.48)	(0.46)	
Uses machinery and/or animal force (1=yes)	0.05	0.07	0.03	0.00***	0.03	0.03	0.00***
	(0.21)	(0.26)	(0.18)		(0.18)	(0.18)	
Value of harvested produce (in Ksh 1,000)	201.18	126.08	240.95	0.11	278.30	208.63	0.20
	(911.44)	(1228.09)	(685.51)		(867.85)	(473.97)	
Production of French beans (in 1,000 kg)	0.75	0.54	0.85	0.40	0.74	0.98	0.59
	(5.55)	(1.20)	(6.66)		(3.60)	(8.92)	
Production of baby corn (in kg)	4.10	10.70	1.06	0.13	1.20	0.90	0.31
	(94.60)	(167.01)	(14.95)		(13.56)	(16.39)	
Total spent in marketing (in Ksh 1,000)	1.30	0.88	1.53	0.16	1.75	1.34	0.28
	(5.95)	(3.64)	(6.91)		(7.89)	(5.95)	
Use of Inputs	0.98	0.98	0.97	0.21	0.97	0.97	0.45
	(0.15)	(0.12)	(0.17)		(0.17)	(0.16)	

Appendix Table 5: Intervention participation factors, AGK table 3 reproduction

	OLS				Probit			
	All (1)	Credit (2)	No Credit (3)	All (4)	All (5)	Credit (6)	No Credit (7)	All (8)
Treatment group included credit	0.125			0.125	0.385			0.125
	[0.079]			[0.080]	[0.227]*			[0.070]*
Member								
Age of member	0.002	0.003	0.001	0.002	0.005	0.003	0.001	0.002
	[0.002]	[0.004]	[0.002]	[0.002]	[0.007]	[0.004]	[0.002]	[0.002]
Literacy	0.025	0.090	-0.031	0.018	0.090	0.095	-0.033	0.020
	[0.059]	[0.130]	[0.045]	[0.061]	[0.185]	[0.120]	[0.049]	[0.062]
Risk tolerance	-0.063	-0.048	-0.080	-0.063	-0.224	-0.046	-0.095	-0.073
	[0.055]	[0.078]	[0.088]	[0.055]	[0.169]	[0.077]	[0.085]	[0.053]
Months as member in SHG	0.001	0.002	-0.000	0.001	0.002	0.002	-0.000	0.001
	[0.001]	[0.001]	[0.001]	[0.001]	[0.003]	[0.001]	[0.001]	[0.001]
Member of SHG is an officer (1=yes)	0.291	0.317	0.224	0.295	0.830	0.302	0.196	0.270
	[0.066]***	[0.100]***	[0.086]**	[0.064]***	[0.172]***	[0.084]***	[0.057]***	[0.049]***
Deposit in a formal bank (1=yes)	0.039	0.068	0.026	0.035	0.141	0.071	0.035	0.043
	[0.041]	[0.082]	[0.029]	[0.042]	[0.123]	[0.080]	[0.029]	[0.041]
Log of total annual household income	0.016	-0.008	0.033	0.126	0.054	-0.010	0.037	0.158
	[0.025]	[0.048]	[0.016]*	[0.082]	[0.075]	[0.044]	[0.016]**	[0.106]
Log of total annual household income squared				-0.016				-0.020
				[0.011]				[0.013]
Number of household members	0.027	0.030	0.021	0.028	0.084	0.030	0.022	0.028
	[0.009]***	[0.017]	[0.006]***	[0.010]***	[0.029]***	[0.017]*	[0.005]***	[0.009]***
Land								
Harvest yield per acre (in Ksh 100,000)	-0.004	-0.006	0.003	-0.004	-0.012	-0.007	0.004	-0.004
	[0.004]	[0.003]*	[0.029]	[0.005]	[0.017]	[0.006]	[0.024]	[0.005]
Proportion of land that is irrigated	0.105	0.101	0.135	0.110	0.334	0.087	0.130	0.110
	[0.081]	[0.151]	[0.084]	[0.079]	[0.238]	[0.144]	[0.067]*	[0.074]
Total landholdings (acres)	0.012	-0.008	0.026	0.014	0.037	-0.008	0.022	0.014
	[0.012]	[0.019]	[0.011]**	[0.013]	[0.034]	[0.018]	[0.010]**	[0.012]
Production								
Grows export crops (1=yes)	0.107	0.060	0.136	0.099	0.331	0.064	0.122	0.097
	[0.060]*	[0.135]	[0.029]***	[0.061]	[0.181]*	[0.128]	[0.025]***	[0.058]*
Sells to market (1=yes)	-0.128	-0.155	-0.106	-0.131	-0.420	-0.157	-0.108	-0.139
	[0.045]***	[0.078]*	[0.043]**	[0.045]***	[0.138]***	[0.074]**	[0.039]***	[0.042]***
Uses hired labour (1=yes)	-0.056	-0.078	-0.015	-0.059	-0.171	-0.081	-0.018	-0.059
	[0.057]	[0.088]	[0.092]	[0.057]	[0.168]	[0.084]	[0.087]	[0.056]
Uses machinery and/or animal force (1=yes)	-0.164	-0.176	-0.091	-0.169	-0.534	-0.172	-0.104	-0.181
	[0.097]	[0.120]	[0.095]	[0.096]*	[0.332]	[0.137]	[0.108]	[0.105]*
Mean dependent variable	0.351	0.445	0.270	0.351	0.351	1.000	0.000	0.538
Number of observations	433	200	233	433	433	200	233	433
R-squared	0.158	0.138	0.167	0.162				

Appendix Table 6: OLS intervention impacts, AGK table 4 reproduction

	Export Crop (1)	Proportion of Land Devoted to Cash Crops (2)	Use of Inputs (3)	Production of French Beans (1,000 kg) (4)	Production of Baby Corn (kg) (5)	Value of Harvested Produce (in Ksh 1,000) (6)	Total Spent in Marketing (7)	Logarithm of Household Income (8)	Loan from Formal Institutions (9)	Deposit in Formal Institutions (10)
Post	-0.005	-0.096	0.009	0.345	21.455	-92.039	0.801	-0.253	-0.047	0.075
	[0.053]	[0.014]***	[0.013]	[1.473]	[19.496]	[114.018]	[0.791]	[0.099]**	[0.017]**	[0.032]**
Post x Treatment	0.246	0.061	0.014	2.232	86.465	138.950	-1.825	0.176	0.032	0.082
	[0.067]***	[0.017]***	[0.016]	[2.867]	[41.947]**	[229.956]	[0.943]*	[0.113]	[0.020]	[0.038]**
Number of observations	1983	1847	1847	1983	1983	1647	1674	1581	1674	1665
R-squared	0.20	0.16	0.05	0.03	0.02	0.03	0.03	0.16	0.05	0.15
Mean of dependent variable	0.563	0.515	0.985	2.034	43.872	228.535	1.053	3.498	0.027	0.789

Note: As mentioned in the text, we did not replicate the IV portion of the results reported in the original publication.

Appendix Table 7: OLS intervention impact by treatment arm, AGK table 5 reproduction

	Export Crop	Proportion of Land Devoted to Cash Crops	Use of Inputs	Production of French Beans (in 1,000 kg)	Production of Baby Corn (in kg)	Value of Harvested Produce (in Ksh 1,000)	Total Spent in Marketing (in Ksh 1,000)	Logarithm of Household Income	Loan from Formal Institutions	Deposit in Formal Institutions
Post	-0.005	-0.096	0.009	0.334	21.591	-89.598	0.801	-0.253	-0.047	0.075
	[0.053]	[0.014]***	[0.013]	[1.476]	[19.464]	[113.408]	[0.792]	[0.099]**	[0.017]***	[0.032]**
Post X credit	0.228	0.078	0.020	-1.172	127.722	374.234	-1.903	0.107	0.014	0.078
	[0.067]***	[0.020]***	[0.018]	[1.138]	[67.326]*	[386.552]	[1.173]	[0.114]	[0.027]	[0.046]*
Post X no credit	0.266	0.038	0.006	6.059	40.080	-87.675	-1.748	0.250	0.050	0.085
	[0.092]**	[0.024]	[0.018]	[5.472]	[27.441]	[106.579]	[0.880]*	[0.133]*	[0.019]**	[0.039]**
Number of observations	1983	1847	1847	1983	1983	1647	1674	1581	1674	1665
R-squared	0.20	0.16	0.06	0.03	0.02	0.03	0.03	0.16	0.05	0.15
p-value of F-test post x credit = post x no credit										
	0.657	0.145	0.465	0.185	0.199	0.197	0.857	0.181	0.103	0.878

Note: As mentioned in the text, we did not replicate the IV portion of the results reported in the original publication.

Appendix Table 8A: Intervention impact by production OLS, AGK table 6 reproduction

	Proportion of Land Devoted to Cash Crops		Use of Inputs		Production of French Beans (1,000 kg)		Production of Baby Corn (kg)	
Grows Export Crops at Baseline	Yes	No	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.117	-0.059	-0.006	0.032	1.514	-1.569	13.235	40.277
	[0.020]***	[0.022]**	[0.007]	[0.029]	[1.849]	[2.426]	[19.918]	[23.447]*
Post x Treat	-0.016	0.104	0.008	0.007	-2.229	6.427	150.759	30.338
	[0.029]	[0.029]***	[0.006]	[0.031]	[1.958]	[4.825]	[84.862]*	[33.668]
Number of observations	895	940	896	939	957	1014	957	1014
R-squared	0.19	0.15	0.04	0.09	0.05	0.04	0.03	0.05
Mean dependent variable	0.601	0.432	0.998	0.973	1.593	2.471	58.274	29.971

Appendix Table 8B: Intervention impact by production OLS, AGK table 6 reproduction

	Value of Harvested Produce (in Ksh 1,000)		Total Spent in Marketing (in Ksh 1,000)		Logarithm of Household Income		Loan from Formal Institutions		Deposit in Formal Institutions	
Grows Export Crops at Baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Post	19.088	-242.484	0.698	1.512	-0.143	-0.364	-0.067	-0.005	0.074	0.085
	[142.153]	[212.881]	[0.966]	[2.112]	[0.093]	[0.223]	[0.016]***	[0.030]	[0.032]**	[0.063]
Post x Treat	-107.578	461.017	-2.037	-2.270	-0.077	0.494	0.042	-0.006	0.045	0.116
	[224.379]	[506.278]	[1.132]*	[2.481]	[0.133]	[0.234]**	[0.021]*	[0.032]	[0.048]	[0.069]
Number of observations	856	779	883	779	842	727	883	779	875	778
R-squared	0.04	0.07	0.03	0.08	0.16	0.2	0.06	0.08	0.14	0.21
Mean dependent variable	221.453	239.833	1.085	1.006	3.603	3.374	0.027	0.026	0.803	0.772

Note: As mentioned in the text, we did not replicate the IV portion of the results reported in the original publication.

Appendix Table 9A: Intervention impact by production history and arm OLS, AGK table 7 reproduction

	Proportion of Land Devoted to Cash Crops		Use of Inputs		Production of French Beans (1,000 kg)		Production of Baby Corn (kg)		Value of Harvested Produce (in Ksh 1,000)	
Grows Export Crops at Baseline	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)	(11)	(12)
Post	-0.117	-0.059	-0.006	0.032	1.516	-1.547	13.632	40.193	20.499	-231.456
	[0.020]***	[0.022]**	[0.007]	[0.029]	[1.850]	[2.405]	[19.942]	[23.342]*	[142.268]	[202.335]
Post x Credit	-0.009	0.141	0.011	0.018	-1.957	-0.005	206.118	55.416	-6.891	1042.071
	[0.032]	[0.033]***	[0.008]	[0.035]	[1.835]	[0.595]	[134.582]	[57.241]	[329.242]	[1130.266]
Post x No Credit	-0.026	0.060	0.005	-0.005	-2.603	12.559	74.770	6.427	-241.193	86.350
	[0.046]	[0.034]*	[0.006]	[0.034]	[2.603]	[8.864]	[48.691]	[25.080]	[162.033]	[118.288]
Number of observations	895	940	896	939	957	1014	957	1014	856	779
R-squared	0.19	0.16	0.04	0.09	0.05	0.04	0.04	0.05	0.04	0.07
Mean dependent variable	0.601	0.432	0.998	0.973	1.593	2.471	58.274	29.971	221.453	239.833
p-value of F-test post x credit = post x no credit										
	0.735	0.038	0.298	0.438	0.745	0.156	0.356	0.395	0.461	0.370

Appendix Table 9B: Intervention impact by production history and arm OLS, AGK table 7 reproduction

	Total Spent in Marketing (in Ksh 1,000)		Logarithm of Household Income		Loan from Formal Institutions		Deposit in Formal Institutions	
Grows Export Crops at Baseline	Yes	No	Yes	No	Yes	No	Yes	No
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Post	0.698	1.515	-0.143	-0.366	-0.067	-0.005	0.073	0.086
	[0.967]	[2.113]	[0.093]	[0.224]	[0.016]***	[0.030]	[0.032]**	[0.063]
Post x Credit	-2.061	-2.096	-0.062	0.330	0.027	-0.035	0.010	0.171
	[1.317]	[2.848]	[0.135]	[0.247]	[0.027]	[0.037]	[0.061]	[0.070]**
Post x No Credit	-2.006	-2.382	-0.099	0.620	0.062	0.013	0.092	0.081
	[1.166]*	[2.371]	[0.204]	[0.250]**	[0.020]***	[0.031]	[0.052]*	[0.073]
Number of observations	883	779	842	727	883	779	875	778
R-squared	0.03	0.08	0.16	0.2	0.07	0.09	0.14	0.22
Mean dependent variable	1.085	1.006	3.603	3.374	0.027	0.026	0.803	0.772
p-value of F-test post x credit = post x no credit								
	0.960	0.834	0.858	0.059	0.147	0.072	0.218	0.112

Note: As mentioned in the text, we did not replicate the IV portion of the results reported in the original publication.

Appendix Table 10: Intervention impact on prices OLS, AGK table 8 reproduction

	All Crops		French Beans		Bananas		Maize (dry)		Beans		Coffee	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post	5.58	5.21	-5.87	-5.87	-2.32	-2.33	-0.12	-0.13	0.14	0.14	0.88	0.87
	[14.62]	[14.68]	[2.80]**	[2.80]**	[0.64]***	[0.64]***	[0.08]	[0.09]	[0.33]	[0.33]	[0.75]	[0.75]
Treatment x	-21.48		-5.34		-0.34		0.09		-0.31		-1.78	
post	[14.64]		[3.06]*		[0.30]		[0.20]		[0.34]		[1.19]	
Credit x		-39.37		-5.24		-0.58		-0.09		-0.36		-1.35
post		[12.36]***		[3.20]		[0.26]**		[0.09]		[0.38]		[1.35]
No Credit x		5.80		-5.55		0.00		0.30		-0.25		-2.49
post		[21.83]		[4.27]		[0.55]		[0.39]		[0.36]		[1.29]*
Observations	1373	1373	729	729	690	690	324	324	265	265	858	858
R-squared	0.03	0.04	0.05	0.05	0.05	0.05	0.03	0.04	0.07	0.07	0.05	0.06
Mean dependent variable	46.30	46.30	18.50	18.50	1.86	1.86	0.19	0.19	0.42	0.42	4.88	4.88
p-value of test post x credit = post x no credit												
		0.02		0.94		0.30		0.34		0.74		0.38

Note: As mentioned in the text, we did not replicate the IV portion of the results reported in the original publication.

References

Ashraf, N, Giné, X and Karlan, D, 2008. Harvard Business School. NOM Working Paper No. 08-065. 'Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya'.

Ashraf, N, Giné, X and Karlan, D, 2009a. Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya. *American Journal of Agricultural Economics*, 91(4), pp. 973–990.

Ashraf, N, Giné, X and Karlan, D, 2009b. Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya. *American Journal of Agricultural Economics*, Appendix, pp. 1–8.

Beegle, K, Carletto, C and Himelein, K, 2012. Reliability of recall in agricultural data. *Journal of Development Economics*, 98(1), pp. 34–41.

Brown, A, Cameron, C and Wood, B, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness*, 6(3), pp. 215–235.

Bruhn, B and McKenzie, D, 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 2009*, 1(4), pp. 200–232.

Byerlee, D, Garcia, A, Giertz, A, Palmade, V and Gurcanlar, T, 2013. *Growing Africa: Unlocking the potential of agribusiness*. Technical report, World Bank. AFTFP/AFTAI.

Carletto, C, Kilic, T and Kirk, A, 2011. Nontraditional crops, traditional constraints: The long-term welfare impacts of export crop adoption among Guatemalan smallholders. *Agricultural Economics*, 42(supplement), pp. 61–75.

Crawford, I, 1997. *Marketing Research and Information Systems*. Food and Agriculture Organisation of the United Nations.

Diskin, P, 1999. *Agricultural Productivity Indicators Measurement Guide*. Food Security and Nutrition Monitoring (IMPACT) Project, ISTI, for the U.S. Agency for International Development.

Djimeu, E and Houndolo, D, 2015. Power calculation for causal inference in social science: Sample size determination and minimum detectable effect. International Initiative for Impact Evaluation mimeo.

Economic Research Service, U.S. Department of Agriculture, 1992. *Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products*. Agricultural Handbook No. 697.

Fan, S, Brzeska, J, Keyzer, M and Halsema, A, 2013. *From Subsistence to Profit: Transforming Smallholder Farms*. International Food Policy Research Institute (IFPRI) Food Policy Report.

Fan, S, Brzeska, J and Olofinbiyi, T, 2015. *2014–2015 Global Food Policy Report*. IFPRI Food Policy Report.

Fermont, A and Benson, T, 2011. *Estimating Yield of Food Crops Grown by Smallholder Farmers: A Review in the Uganda Context*. IFPRI Discussion Paper 01097.

Hoenig, J and Heisey, D, 2001. The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. *The American Statistician*, 55, pp. 19–24.

Kherallah, M, Delgado, C, Gabre-Madhin, E, Minot, N and Johnson, M, 2002. *Reforming Agriculture Markets in Africa*. IFPRI.

Lenth, R, 2007. *Post Hoc Power: Tables and Commentary*. The University of Iowa. Department of Statistics and Actuarial Science, Technical Report No. 378.

Obare, GA, 2000. *The impact of road infrastructure on input use and farm level productivity in Nakuru District, Kenya*. PhD thesis, Egerton University.

Stephens, MA, 1992. *An Appreciation of Kolmogorov's 1933 Paper*. Technical Report, Stanford University.

Strasberg, P, Jayne, T, Yamano, T, Nyoro, J, Karanja, D and Strauss, J, 1999. *Effects of Agricultural Commercialization on Food Crop Input Use and Productivity in Kenya*. Food Security International Development Working Papers 54675.

Verduzco-Gallo, I, Ecker, O and Pauw, C, 2014. *Changes in food and nutrition security in Malawi: Analysis of recent survey evidence*. IFPRI Malawi Strategy Support Programme: Working Paper No. 6.

Wiggins, S, Argwings-Kodhek, G, Leavy, J and Poulton, C, 2011. *Small farm commercialisation in Africa: Reviewing the issues*. Future Agricultures, Research Paper 023.

Wood, B and Djimeu, E, 2014. Requiring fuel gauges: A pitch for justifying impact evaluation sample size assumptions. *3ie Evidence Matters blog*. Available at: http://blogs.3ieimpact.org/requiring-fuel-gauges-a-pitch-for-justifying-impact-evaluation-sample-size-assumptions/ [Accessed 14 May 2015].

Wood, B, Nelson, C, Kilic, T and Murray, S, 2013. *Up in smoke?: Agricultural commercialization, rising food prices and stunting in Malawi*. World Bank Policy Research Working Paper No. 6650.

World Bank. Note on Conversion factors for food item-non-standard measurement unit combinations in the Malawi third integrated household survey (IHS3) 2010/11 Data. Available at: http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1233781970982/5800988-1271185595871/Malawi_IHS3_Food_Item_Conversion_Factors.pdf [Accessed 14 May 2015].

Publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/:

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Replication Paper Series

International Initiative for Impact Evaluation
1625 Massachusetts Ave., NW
Suite 450
Washington, DC 20036
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org