

Household welfare measurement in Bangladesh

A Tale of Two Short Consumption Modules

Luisa Natali and Chris de Neubourg

Office of Research Working Paper

WP-2014-No. 17 | December 2014

INNOCENTI WORKING PAPERS

UNICEF Office of Research Working Papers are intended to disseminate initial research contributions within the programme of work, addressing social, economic and institutional aspects of the realization of the human rights of children.

The findings, interpretations and conclusions expressed in this paper are those of the authors and do not necessarily reflect the policies or views of UNICEF.

This paper has been extensively peer reviewed both internally and externally.

The text has not been edited to official publications standards and UNICEF accepts no responsibility for errors.

Extracts from this publication may be freely reproduced with due acknowledgement. Requests to utilize larger portions or the full publication should be addressed to the Communication Unit at florence@unicef.org.

For readers wishing to cite this document we suggest the following form:

Natali, L. and C. de Neubourg (2014). Household Welfare Measurement in Bangladesh: A tale of two short consumption modules, *Innocenti Working Paper* No. 2014-17, UNICEF Office of Research, Florence.

A Methodological Annex accompanies this paper and is available on the Office of Research website at: www.unicef-irc.org/publicationis/766

© 2014 United Nations Children's Fund (UNICEF)

ISSN: 1014-7837

THE UNICEF OFFICE OF RESEARCH

In 1988 the United Nations Children's Fund (UNICEF) established a research centre to support its advocacy for children worldwide and to identify and research current and future areas of UNICEF's work. The prime objectives of the Office of Research are to improve international understanding of issues relating to children's rights and to help facilitate full implementation of the Convention on the Rights of the Child in developing, middle-income and industrialized countries.

The Office aims to set out a comprehensive framework for research and knowledge within the organization, in support of its global programmes and policies. Through strengthening research partnerships with leading academic institutions and development networks in both the North and South, the Office seeks to leverage additional resources and influence in support of efforts towards policy reform in favour of children.

Publications produced by the Office are contributions to a global debate on children and child rights issues and include a wide range of opinions. For that reason, some publications may not necessarily reflect UNICEF policies or approaches on some topics. The views expressed are those of the authors and/or editors and are published in order to stimulate further dialogue on child rights.

The Office collaborates with its host institution in Florence, the Istituto degli Innocenti, in selected areas of work. Core funding is provided by the Government of Italy, while financial support for specific projects is also provided by other governments, international institutions and private sources, including UNICEF National Committees.

For further information and to download or order this and other publications, please visit the website at www.unicef-irc.org.

Correspondence should be addressed to:

UNICEF Office of Research - Innocenti
Piazza SS. Annunziata, 12
50122 Florence, Italy
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

HOUSEHOLD WELFARE MEASUREMENT IN BANGLADESH: A TALE OF TWO SHORT CONSUMPTION MODULES

Luisa Natali¹ (lnatali@unicef.org; luisanatali@gmail.com) and Chris de Neubourg²

¹ UNICEF Office of Research

² TIAS School for Business and Society; Tilburg University

Abstract. Two short consumption modules were piloted in Bogra and Sirajganj (Bangladesh) in May-June 2012 as part of the Global MICS5 Pilot. This paper aims at validating this exercise and assessing the accuracy and reliability of the consumption estimates obtained. The use of a benchmark consumption module is essential in order to assess how well the two short options fare; the analysis therefore consists of a systematic comparison of both short modules with a benchmark. The attempt made is to isolate and test the impact of the length (degree of commodity) of the consumption questionnaire on the quality of consumption and poverty estimates as well as distributional measures obtained. We conclude that it is feasible to include a short consumption module in MICS (Multiple Indicator Cluster Surveys). The Bangladesh experience suggests that this module can give accurate predictions of aggregate consumption and poverty, allowing for the analysis of monetary and non-monetary dimensions of welfare together. However the module cannot be used to analyze individual consumption groups (like food, non-foods, etc.) or consumption patterns.

Keywords: consumption; expenditure; poverty; survey design; Bangladesh; short module; pilot.

JEL classification: C83, I32, D12

Acknowledgements: We are extremely grateful to: Bruno Martorano for his continuous support, for reviewing the paper and for providing useful comments – his help has been invaluable; to Attila Hancioglu, Bo Pedersen and the Global MICS Team for their assistance and for making it possible to test the two short consumption modules in Bangladesh; to the Bangladesh Bureau of Statistics (BBS) and in particular to Shaheen Mohammed and Dipankar Roy; to Stuart Cameron, Mahruf Shohel, Andrea Verdasco for their help before and after travelling to Bangladesh; to Martin Evans for providing comments; to Michelle Seroussi and Thi Minh-Phuong Ngo for helpful discussions; to Kenneth R. Simler for providing the code for simulating poverty from predicting consumption; to Faizuddin Ahmed and finally to Sudhanshu Handa for his supervision and patient help. The authors would like to thank Sander den Boer, Lucian Mircescu and Marta Moratti who worked on the compilation of the short consumption modules during the first phase of the project.

TABLE OF CONTENTS

1. Introduction	6
2. Literature review: short versus long consumption modules	7
3. Experimental setting and data description	8
4. Compute comparable consumption aggregates	12
5. Descriptive statistics: are the three sub-samples identical in all respects?	12
6. Data analysis and findings	15
6.1 Consumption levels	15
6.2 Poverty estimate	32
6.3 Distributional estimates	36
7. Robustness checks and limitations	43
8. Conclusions	43
References	45

The accompanying Methodological Annex is available at: www.unicef-irc.org/publications/766

1. INTRODUCTION

Through the Multiple Indicator Cluster Survey (MICS) programme UNICEF supports the collection of data on the situation of children and women. MICS are multi-topic surveys that cover different areas such as education, health, gender equality, rights and protection. MICS provide a wealth of data but they lack information on income and/or consumption; indeed, wealth indices represent the only way to investigate distributional aspects (Howe et al., 2008).

The inclusion of a monetary measure within MICS is therefore potentially extremely advantageous. First, it would make it possible to quantify a household's current levels of monetary welfare or poverty by providing a measure of consumption levels and poverty rates; it would be an absolute rather than a relative measure of welfare, in contrast to the wealth index – where a household's wealth is measured relative to other households in the sample rather than to a set standard. Second, it would make it possible to investigate more precisely distributional issues; third, it would allow for welfare comparisons across countries and time. Finally, and most importantly, if it is true that MICS surveys often remain under-utilized in research due to the lack of a monetary measure, its inclusion could significantly increase MICS usage for policy analysis and research purposes.

However information on income, consumption and expenditure are complex and time-consuming to collect and MICS is already a long household survey. For the inclusion of such information within MICS, it is therefore important that the module does not represent a substantial burden by significantly expanding the length of the questionnaire and, at the same time, it should not jeopardize the existing MICS survey design.¹ Taking these two constraints into account and building on existing literature, two short consumption modules were developed by the UNICEF Office of Research at Innocenti in collaboration with the Global MICS Team (SMS, NY). The two different short consumption modules were then piloted in Bogra and Sirajganj (Bangladesh) in May-June 2012 as part of the Global MICS5 Pilot.

This paper aims to validate this exercise and assess the accuracy and reliability of the consumption estimates obtained. The use of a benchmark consumption module is essential in order to assess how well the two short options fare; the analysis therefore consists of a systematic comparison of both short modules with a benchmark. The attempt made is to isolate and test the impact of the length (degree of commodity) of the consumption questionnaire on the quality of consumption and poverty estimates as well as distributional measures obtained. Analysing the pilot data and understanding if and how much the short consumption modules estimates differ from the benchmark is the first step towards understanding whether the inclusion of the tested short consumption modules within MICS is possible and meaningful.

The paper is structured as follows. *Section 2* briefly reviews the literature on short versus long consumption modules. *Section 3* provides a description of the pilot and data. The consumption aggregates are then computed in *Section 4* while some preliminary descriptive statistics are reported in *Section 5*. *Section 6* describes the data analysis carried out and discusses the results for consumption estimates (sub-section 6.1), poverty rates (sub-section 6.2) and distributional measures (sub-section 6.3); as the samples under review are non-equivalent, several

¹ i.e. sampling, number of visits by interviewer, main respondent and so on.

methodologies are applied. *Section 7* describes the procedure for checking the robustness of the results and highlights some limitations. Finally, *Section 8* concludes.

2. LITERATURE REVIEW: SHORT VERSUS LONG CONSUMPTION MODULES²

While different indicators exist to perform welfare analysis, for a long time consumption has been favoured by economists as a proxy for living standards. However, consumption measurement is a challenging and time-consuming task as interviewers need to refer to a long list of food and non-food items to collect data. The degree of commodity detail is therefore an important determinant of the duration time of long and detailed consumption modules that can cover over 400 items. Therefore, short consumption modules – that attempt to reduce the length of the interview and respondents' fatigue³ – could have enormous potential advantages. For this reason, a strand of the literature has compared short expenditures and/or consumption expenditures modules with longer and more detailed ones in order to test the accuracy of the resulting estimates.

Available evidence, as reviewed in Moratti and Natali (2012), seems to indicate that short modules underestimate consumption with respect to longer ones resulting in lower levels of recorded consumption and higher poverty rates. If shorter modules tend to bias consumption downwards, non-food consumption is usually underestimated more than food consumption. There is also some indication that 'the fraction by which consumption is underestimated increases as consumption rises' (Pradhan, 2001:22).

However, one of the most complete, recent and authoritative studies in the field (Beegle et al., 2010) finds that short modules may actually result in a smaller downward bias compared to the benchmark than other longer consumption modules. In particular, they find that the 'subset' module, which includes 17 food items that constitute, on average, 77 per cent of food consumption expenditure, "performs very close to the longer list form and may be a suitable substitute to longer list recall modules" (Beegle et al., 2010:31). Also, the 'subset' module performs much better than the alternative short module, defined as a 'collapsed' module, where the 58 food items from a longer module are aggregated into 11 comprehensive categories. This result has important implications for the design of the short consumption module as it suggests that 'subset' modules, which identify the most common items consumed, provide more reliable information than modules that use item categories. The literature on the impact of short versus long consumption modules on distributionally-sensitive measures is scant. However, results from rigorous studies indicate that short consumption modules are good at capturing the relative ranking of households (as reviewed in Moratti and Natali, 2012).

A critical review of the available evidence points to a number of factors that hinder the ability to draw firm conclusions. First, experiments have been scarce and mostly run in single countries in different regions at different points in time. Second, different methodologies have been adopted. Third, there is a difference – often overlooked in the literature – between subset and collapsed questionnaires; the latter asks about consumption in a number of comprehensive broad categories, while the former uses a reduced number of single items (not comprehensive but expected to mirror total consumption). More generally, in field experiments short and long

² A more extensive literature review is provided by Moratti and Natali (2012).

³ As well as non-response.

consumption modules do not always capture the same comprehensive definition of consumption. Fourth, the use of different recall periods between the modules tested acts as a confounding factor. Last, the use of different benchmark modules constitutes another confounding factor in these studies: sometimes personal diaries are used, other times long recall questionnaires and so on; studies rarely provide a rationale on how the long questionnaire has been or should be selected as a benchmark and seem to implicitly assume that any long questionnaire is closer to actual consumption.

The review of the literature, and the above-mentioned factors in particular, indicates that there is room for further investigation. In particular, future field experiments should compare short consumption modules – clearly distinguishing between collapsed and subset modules – to a well-defined long benchmark, chosen for its ability to produce consumption estimates that are the closest to actual consumption. Survey design should be the same across the short and long consumption modules: the method of data capture, the level of respondent, as well as the length of the reference period, should be consistent in order to isolate the impact of the length or level of detail in consumption questionnaires on the accuracy of the resulting estimates. Several/multiple methodologies (quantitative analysis) should be applied and, where funding allows, different contexts should be explored (a multiple country comparative analysis).

3. EXPERIMENTAL SETTING AND DATA DESCRIPTION

The experiment was carried out as part of the Global MICSS Pilot in May-June 2012 in the Bogra and Sirajganj districts of Bangladesh (Rajshahi division). The pilot was hosted by the Government of Bangladesh's Bureau of Statistics (BBS) and the UNICEF Country Office; it was preceded by a month-long intensive training provided mainly by the Global MICS Team as well as a number of international experts. The trainees – around 50 – were interviewers, supervisors, editors, measurers and data entry operators recruited directly from BBS.

The aim of the Global MICSS Pilot was to test whether the new MICS survey instruments worked properly in order to compile the final MICS questionnaires to be used for the 5th round of MICS surveys during the next 3 years in at least 40 countries. Four questionnaires were tested: 1) the household questionnaire; 2) the questionnaire for individual women; 3) the questionnaire for individual men; and 4) the questionnaire for children under 5. Each questionnaire encompasses several modules. Within these questionnaires, some of the modules tested were new (children left behind; water testing; and short consumption modules) whereas some were slightly modified and/or enhanced (child discipline; child labour; immunization schedule; contraception; care of illness). The pilot was also used to test 'processes'; in particular, two different ways of collecting data (paper versus PDAs – personal digital assistants) and a new anthropometry training were tested.

As the overall goal of the pilot was to test the tools and not the results, the Global MICS team decided not to pursue a representative sample.⁴ However, in order to be able to recreate the tables from the MICS tabulation plan, a sufficient number of observations were needed. The sample size comprised 731 households interviewed using the traditional paper (715 completed) method. The households were drawn from two districts – Bogra and Sirajganj – in the Rajshahi

⁴ Moreover, no updated sample enumeration areas were available.

division, for two main reasons: first, an attempt to cover one of poorest districts, namely Sirajganj; and second, for practical reasons, namely the proximity to the training base in Bogra.

During the field experiment two different short consumption modules were administered: 1) the 'itemized' (or 'subset') consumption module; and 2) the 'categorized' (or 'collapsed') consumption module.⁵ Table 1 reports the more salient characteristics of the 2 modules.

Table 1 – Description of consumption modules

	Categorized (or collapsed)		Itemized (or subset)		HIES (Household Income and Expenditure Survey)	
Length module	Short		Short(est)		Long (benchmark)	
Food content	<i>Food consumption <u>inside</u> the household.</i> Food items are aggregated into 14 mutually exclusive broad comprehensive categories. Info on value (not quantity) is collected.	purchases (7day) home production (last 12 months) gifts and/or assistance (last 12 months)	<i>Food consumption <u>inside</u> the household.</i> 15 food items ¹ . Info on value and quantity is collected.	purchases (7day) home production (last 12 months) gifts and/or assistance (last 12 months)	High level of disaggregation - around 140 items - including both <i>food consumption inside and outside the household</i> . Data collected daily (or weekly) for 14 days (2 weeks). Quantity and value collected. Major source of consumption for each item indicated as either 1-purchase, 2-wage in kind, 3-self-production, 4-gift	
Non-food content	12 types of <i>non-food expenditures</i> . No information on durables is collected. Only value (no quantity) is collected.	Education, health and extraordinary expenditures (1 year recall). Other non-food expenditures (1 month recall) (imputed/) rent (reference period preferred by respondent)	7 types of <i>non-food expenditures</i> . No information on durables is collected. Only value (no quantity) is collected.	Education, health and extraordinary expenditures (1 year recall). Other non-food expenditures (1 month recall) (imputed/) rent (reference period preferred by respondent)	Info on around 215 non-food expenditure items. Information on durables is collected. Only value (no quantity ²) is collected.	
Sample (N hhs)	93		97		1,580 ³	
Survey	Global MICSS Pilot (2012)		Global MICSS Pilot (2012)		Bangladesh HIES (2010)	
Respondent	an adult knowledgeable member		an adult knowledgeable member		Most knowledgeable member	
Unit of analysis	household level		household level		household level	
Extra tools	Visual aid + sheets for interviewers ⁴		none (sheets for interviewers)		none	

¹ The 15 food items included are those that constitute, on average, 77 percent of food consumption expenditure in Bangladesh based on the previous Household Income and Expenditure Survey (2010) and represent therefore the most frequently consumed foods.

² There are however some exceptions; for further details please consult directly the BBS website.

³ Overall, Bangladesh HIES (2010) covers 12,240 households. However, only 440 households from Bogra and Sirajganj (Rajshahi division) are used as a benchmark.

⁴ The sheets for interviewers and 'visual aid' - meant to help interviewers with computations and households with recalling - are reported along with the modules in the Methodological Annex.

⁵ The modules, along with a short description, are attached in Annex 1 of the Methodological Annex.

During the Global MICS5 pilot in Bangladesh both short consumption modules were administered as part of the household questionnaire, meaning that any adult (15 or over) or knowledgeable member in the household was eligible to be the respondent.⁶ The consumption modules were administered only to a subset of households;⁷ 93 households received the categorized module and 97 the itemized one. Although as already mentioned, no sampling strategy was used to allow for comparability across sub-samples, alternate distribution of modules was followed in order to allow for random administration of the itemized and categorized modules.

In order to test and validate data collected, an experimental setting would have been needed. Three equivalent samples should have been provided with the categorized, itemized and benchmark modules respectively to ensure that the differences in the estimated levels of consumption were generated from the differences in the module design and are not due to differences in the sample design. However, only a convenience sampling was carried out and although foreseen in the initial plans and strategy, it was not possible to administer a benchmark consumption module during the pilot;⁸ therefore, in order to assess whether data collected are valid, the Bangladesh 'Household Income and Expenditure Survey' (HIES 2010) is used as a benchmark (See Table 1).

The Bangladesh HIES (2010) is a large national representative survey that sampled 12,240 households over a 12 month period between February 2010 and January 2011. Data are collected by BBS following a two-stage sample design of 612 primary sampling units, from 16 strata (6 rural, 6 urban and 4 statistical metropolitan areas). The survey collected detailed information on a range of socio-economic topics such as education, health, migration, economic activities and wage employment, enterprises (agricultural and non-agricultural), housing, income and assets (including remittances) and, last but not least, consumption.⁹

The original 2010 HIES survey covers each of the 6 divisions of Bangladesh; however, as benchmark we use a subsample of 440 households from Bogra and Sirajganj districts only, based on the assumption that such data is the most comparable to the Global MICS5 pilot sample.¹⁰ Figure 1 shows, in clockwise order, the Rajshahi division, it then highlights the Bogra and Sirajganj districts and finally shows the distribution of households interviewed by module type (categorized and itemized respectively) in the sub-districts covered by the pilot.

⁶ Ideally the most knowledgeable member would be interviewed; this was not possible in order to avoid changes in MICS survey design.

⁷ Each interviewer would administer a consumption module only to the first household interviewed every day of the pilot.

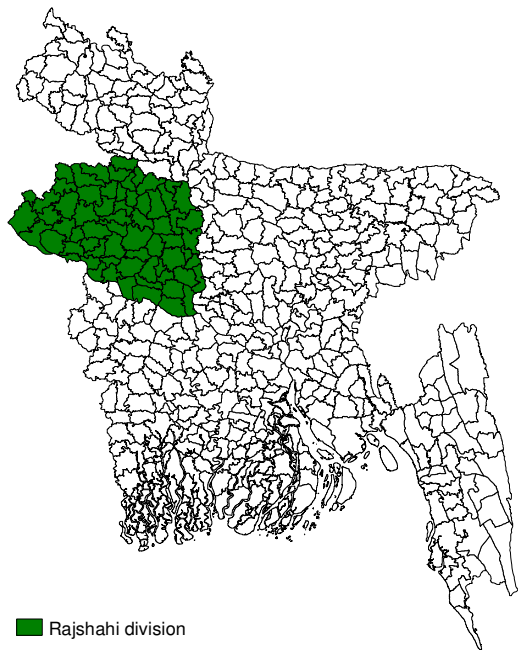
⁸ Ideally, a personal diary with frequent supervision would have been used similarly to Beegle et al. (2010) as this is usually considered the best proxy of true consumption.

⁹ See BBS (2011) for the full Report of HIES 2010.

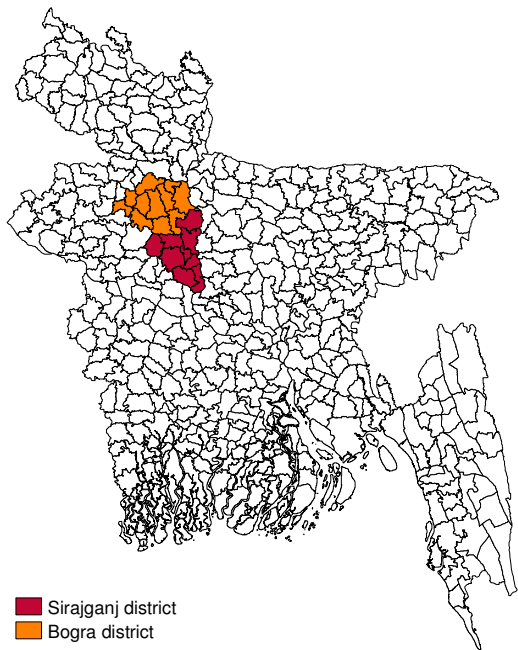
¹⁰ We also carried out the same analysis using as a benchmark the division-level HIES data for Rajshahi division (n=1580) as such data is representative (at the division level). However, adjacent sub-districts are more likely to be similar to the sub-districts sampled and therefore these results are reported throughout the paper.

Figure 1 – Bangladesh Maps

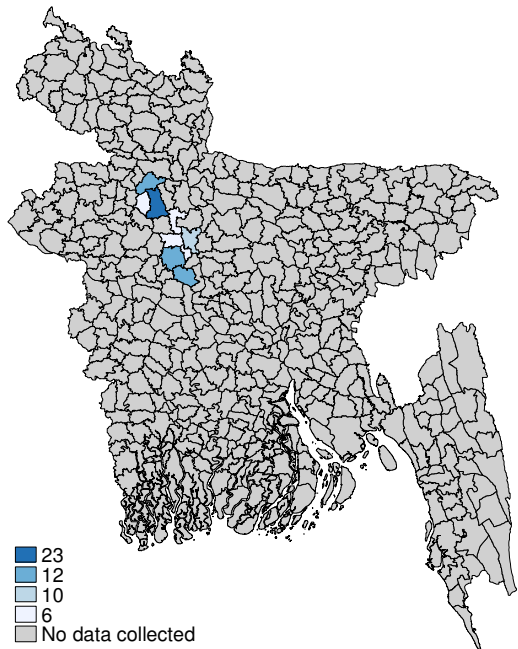
Rajshahi division



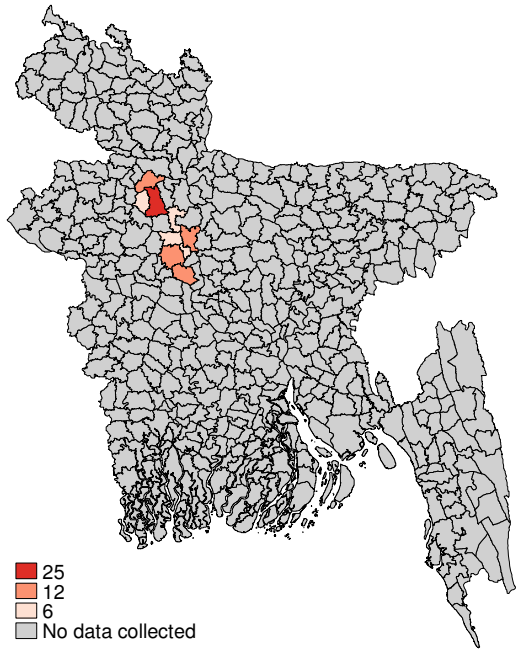
Bogra and Sirajganj districts (Rajshahi division)



Categorized module - distribution of households interviewed (93) in Bogra and Sirajganj districts



Itemized module - distribution of households interviewed (97) in Bogra and Sirajganj districts



4. COMPUTE COMPARABLE CONSUMPTION AGGREGATES

In order to carry out a systematic comparison of the two short modules with the benchmark, it is necessary to construct a comparable consumption aggregate for each of the three samples (or consumption modules). As far as possible, the main existing guidelines have been followed in the computation of the consumption aggregate (Deaton and Grosh, 2000; Deaton and Zaidi, 2002) although the aim is to compute a non-durable consumption aggregate.¹¹

A non-durable consumption aggregate was computed for each module (HIES 2010, categorized and itemized). The main steps are described briefly below; however, a detailed description of how consumption was computed in each questionnaire is provided in the methodological annex (Annex 2).

First, consumption has been expressed in monthly terms. Indeed information on food/non-food expenditures is collected using different recall periods; it is therefore necessary to convert all reported consumption to a common reference period – in this case a monthly value is used. Second, per capita consumption is computed; the consumption aggregate as well as consumption sub-aggregates are divided by household size to get per capita figures.¹² Third, in order to obtain real consumption, consumption estimates were adjusted to account for geographical price differences. Fourth, we checked for outliers; conservative criteria are used to detect and fix outliers. Finally, items to be included in the consumption aggregate are added together. In this paper, food consumption includes both food consumption within the household (from purchases, home production, gifts/assistance) and food consumption outside the household. With respect to non-food consumption, consumer durables data are not collected whereas some lumpy expenditures such as festivities are excluded (although data was collected). Also rent was excluded from the final aggregate as very few households actually reported paying rent.¹³

In the itemized module food expenditures within the household also had to be scaled up in order to compare its estimates with the other modules. Finally, the consumption aggregate from the HIES 2010 was also deflated (inflation over time) to make it comparable to consumption from the pilot survey data that refer to 2012.

5. DESCRIPTIVE STATISTICS: ARE THE THREE SUB-SAMPLES IDENTICAL IN ALL RESPECTS?

Before undertaking the analysis of consumption data, the three samples are compared. Similar samples are necessary to ensure that the differences in the estimated levels of consumption are

¹¹ The two short consumption modules were intentionally developed to capture non-durable consumption. Although guidelines call for a comprehensive definition of consumption, it was decided to drop consumer durables from the short modules. This is driven on one hand by practical reasons as durables inclusion could be demanding in terms of interview time; indeed, when dealing with durable goods (such as home, vehicles, washing machine, computers, etc.) what should be computed is not the expenditure itself but the flow of services that they yield and, in order to compute this flow of services for durable goods, information is needed on the age of each durable good as well as on its original and current value (Deaton and Zaidi, 2002; Deaton and Grosh, 2000). On the other hand, the decision was taken in an attempt to be accurate; in practice, estimating the value of service flows involves crucial assumptions such as definition of durable good, depreciation rate of different items and so on (Deaton and Zaidi, 2002; Deaton and Grosh, 2000); to collect information on consumer durables, it is necessary to ask about a detailed/long list of items that are likely to be reported with substantial error. It was therefore decided to sacrifice the comprehensiveness of the consumption definition by dropping these noisy questions and focus instead on a non-exhaustive list of items believed to be better measured.

¹² This procedure takes into account only the size but not the composition of households. Further robustness checks/sensitivity analysis could involve re-computing consumption using adult equivalent scales.

¹³ There is indeed a trade-off between the inclusion and the quality of data. More discussion is provided in Annex 2.

generated from the differences in the module design and are not due to differences in the sample design.

A test for the similarity of characteristics which are expected to be correlated with consumption is therefore run across the three sub-samples.¹⁴ This test is helpful to establish whether those households who received the short consumption modules are similar to those who received the long questionnaire, namely they are representative of the same population. This is extremely important as no randomized assignment of households to different modules was implemented during the Global MICS5 Pilot.

A number of characteristics such as location of residence, household size and composition, age and gender of head of the household, education and so on, are compared across the three samples. The t-statistics/chi2-statistics test the null hypothesis that the mean values/proportions are the same across the short and long consumption modules.

Table 2 reports significance tests between the benchmark and each of the two short consumption modules (pair-wise t-tests). Given HIES 2010 administered a different household questionnaire to that of the Global MICS5, a reduced number of variables is strictly comparable across samples. However, the table shows that many of these variables are statistically different compared to the reduced HIES sample covering Bogra and Sirajganj only (n=440) used as a benchmark.¹⁵

Table 2 Basic household characteristics by consumption module assignment (HIES, categorized and itemized); pairwise t-tests.

		(1)	(2)	(3)
Part 1 - Means		Categorized	Itemized	HIES - Bogra and Sirajganj districts
Age of the head of the household		45.6	49.6 ***	45.2
Household size		4.0	4.5 **	4.1
Number of women (15-49)		1.1	1.1	1.1
Number of men (15-49)		1.0	1.3 ***	1.0
Number of children under 5		0.3 **	0.4	0.4
Number of cattle ¹		0.6 *	0.7	0.9
Number of goats		0.5	0.5	0.5
Number of sheep		0.1	0.1	0.1
Number of chickens and ducks		2.6	2.9	6.1
Part 2 - Proportions				
Household characteristics	Household with children (<=15 y.o.)	68.8	73.2	74.5
	Male head of the household	89.3	93.8 *	88.0
	Islam - head of the household	88.2 ***	90.7 *	95.7
	Hinduism - head of the household	11.8 ***	9.3 *	4.3
Household has:	Electricity	74.2 ***	75.3 ***	53.6
	Radio	5.4	8.2	6.5
	Television	51.6 ***	51.6 ***	33.1
	Landline ²	2.2	1.0	1.5
	Refrigerator ³	23.7 ***	15.5 ***	5.0

¹⁴ Even if there are differences between the samples, it may not matter unless these are thought to determine the level of the dependent/outcome variable – consumption – or its measurement.

¹⁵ See footnote 9.

<i>Any household member has:</i>	Watch ⁴	39.8		38.1		42.4
	Mobile	85.0	***	81.4	***	57.8
	Bicycle	25.8		37.1	**	25.1
	Motorcycle/scooter	9.7	***	4.1		2.1
	Car/truck ⁵	1.1		0.0	***	0.7
<i>Properties</i>	Dwelling owned	86.0		88.7		86.3
	Land for agriculture	43.0		33.0		38.7
	Animals	53.8	***	56.7	**	70.3
<i>Location</i>	Rural	74.2	***	75.3	***	88.5
<i>Drinking water source:</i>	Tubewell, Borehole ⁶	92.5	***	1.0	***	99.2
	Supply water ⁷	7.5	***	0.0	**	0.2
<i>Roof material</i>	Metal tin roof	91.4		91.8		94.0
	Bamboo roof ⁸	1.1		1.0		1.4
	Cement roof	7.5	***	7.2	***	2.2
<i>Exterior walls material</i>	Brick/cement walls	19.4	***	15.5	**	8.8
<i>Respondent is</i>	Head of the household	29.0	***	22.7	***	46.6
	Wife of the head of the household	59.1	**	60.8	**	46.9
	Other	11.8	*	16.5	***	6.6
<i>Education⁹</i>	None	33.3	***	42.3	***	61.7
	Primary	25.8	***	21.7	*	14.0
	Secondary	31.2	*	28.9		21.7
	Higher	9.7	***	7.2	**	2.6

Sample	93	97	440
--------	----	----	-----

*, **, *** indicate statistical difference in means/proportions across at least two pairs at 10, 5, and 1 percent respectively.

¹ Cattle refer to cattle, milkcows and bulls in MICS and to cattle only in HIES.

² Landline represents owning a non-mobile phone in MICS whereas it means having a landline telephone connection in HIES 2010.

³ In HIES it could be refrigerator or freezer.

A higher proportion of households in the pilot samples (i.e. categorized and itemized) have electricity (74 and 75% in the categorized and itemized samples respectively against 54% in the HIES benchmark), a television (52% in the pilot samples against 33% in the benchmark), a refrigerator (only 5% in the HIES benchmark owns a refrigerator compared to 24% and 16% in the categorized and itemized samples respectively) and a mobile phone (85% and 81% in the categorized and itemized samples respectively against 58% in the HIES benchmark). 'Categorized' households are also almost five times as likely as HIES households to have a motorcycle or scooter (10% against 2%). Moreover, the HIES sample is more rural¹⁶ (88.5% against 74 and 75%) and less educated (indeed, a higher percentage of households has no education, whereas relatively fewer households have some form of education at any level). The fact that samples differ in characteristics relevant to consumption is clearly not ideal for our analysis; it means, in fact, that simply comparing the estimates of interest across the three samples will not be enough to conclude whether the differences in consumption levels between the short modules and the benchmark are driven by the type of module administered and its design rather than by differences in sample characteristics.

¹⁶ The fact that HIES households own more animals most likely reflects the fact that HIES is more rural rather than capturing asset accumulation or welfare.

However, this analysis is helpful in telling us which way to expect the bias. Indeed, from the comparison of basic household characteristics among samples, HIES households seem poorer; we could therefore expect consumption to be higher in the pilot samples with respect to the benchmark sample.

It is also interesting to compare the two pilot samples, i.e. the categorized and the itemized samples. Table A1 in Annex 3 reports significance tests. The two samples are comparable in most household characteristics and it would therefore seem that overall the random assignment worked. However, there are some noticeable differences between the two groups that should be kept in mind when assessing the results later on. At the demographic level, the categorized sample appears to have slightly fewer household members, slightly younger heads of the household and a higher number of working-age men. Every single household in the itemized sample uses tubewell/borehole as the main source of drinking water with no space for more improved sources. Finally, in the categorized sample around 44% of the households report having at least one household member with a bank account whereas in the itemized sample that proportion is much lower at 24%. We could therefore hypothesize that the categorized module is slightly better off than the itemized sample and that its consumption would therefore be higher than consumption reported by households in the itemized sample.

6. DATA ANALYSIS AND FINDINGS

The first part of the data analysis (sub-section 6.1) compares consumption levels; the second part (sub-section 6.2) poverty rates and finally the third (sub-section 6.3) how households rank (and/or distributionally-sensitive measures).

In each of these sub-sections, short consumption modules estimates are first assessed against a benchmark defined as the reduced HIES sample covering Bogra and Sirajganj (n=440). Results from this comparison are only preliminary as it cannot be concluded that differences observed are driven by module rather than sample design. Still, this first part of the analysis is interesting as it provides some first estimates that can be used as a benchmark against which we can compare the estimates obtained later with other methodologies that reduce the existing bias. Indeed in sub-sections 6.1, 6.2 and 6.3, several methodologies are applied in order to take into account the non-equivalence of the samples under review. Our assumption is that any difference observed in this first part of the analysis should be reduced as we attempt to eliminate existing confounding factors. Reducing the bias will help in trying to isolate the impact of module assignment that is the main interest of this paper. Each section concludes with a summary table to recap findings.

6.1 Consumption Levels

This first part of the data analysis compares consumption levels. Consumption estimates from each short module are assessed against a benchmark.

First, consumption means are compared (sub-section 6.1.1); here the benchmark is defined as the reduced HIES sample covering Bogra and Sirajganj (n=440). A first attempt is then made to overcome the problem of comparability that follows from the use of non-equivalent samples by use of a simple OLS consumption regression.

Household total consumption is then predicted from the national HIES 2010 into the two short consumption modules samples and used as a benchmark (sub-section 6.1.2). This methodology is interesting not only as it allows to control for a number of variables correlated with consumption (such as household characteristics, assets owned by the household, quality of housing, access to facilities, etc.) that vary across the three samples but also as it has sometimes been put forward as an alternative to the inclusion of short consumption modules in multi-topic surveys such as MICS which are already lengthy.

Finally, in the last part of the analysis (sub-section 6.1.3), propensity score matching is applied. Given that the samples differ in some relevant household characteristics which can determine either the level of consumption or its measurement, an attempt is made to reduce such bias by matching observations that have similar characteristics.

The section concludes with a summary table of results (sub-section 6.1.4).

6.1.1 Preliminary comparison of estimates across the three samples: differences in mean consumption estimates

The test for the similarity of the three samples tested whether the households in the three subsamples were drawn from the same underlying population. The paper now turns to analyse how the three subsamples vary in terms of their consumption levels.

Before starting our comparative analysis, however, we report the estimates for monthly consumption expenditure per household as reported in the Bangladesh HIES 2010 Report. At the national level, the average monthly consumption expenditure per household was Taka 11,003 in 2010 (around USD 140). The average consumption expenditure was lower in rural areas, at Taka 9,436 per month, and higher in the urban area, at Taka 15,276 (see Table 3). Although we are interested in current levels of consumption, Table 3 also provides estimates for 2005 consumption in order to show the impressive improvement in living standards that was recorded over 5 years (the monthly average consumption increased by 84.5% over the year 2005).

Table 3: Average monthly consumption expenditure per household by residence

Year	Residence	Average <u>household</u> consumption <u>per month</u>	Average household size	Average <u>per capita</u> consumption <u>per month</u>		Average <u>per capita</u> consumption <u>per day</u>
				Taka	US \$	US \$
2010	National	11,003	4.5	2,445	31.3	1
	Rural	9,436	4.53	2,083	26.7	0.9
	Urban	15,276	4.41	3,464	44.4	1.5
2005	National	5,964	4.85	1,230	15.8	0.5
	Rural	5,165	4.89	1,056	13.5	0.4
	Urban	8,315	4.72	1,762	22.6	0.7

Source: BBS (2011). Note: The 2010 HIES Report provides average household consumption per month and household size by residence. Based on these, average per capita consumption reported in this table was computed by author; exchange rate used is as follows: US\$1 = Tk 78.01 (April 2013).

Clearly, monthly consumption expenditures vary greatly by administrative division as shown in Table 4. We are particularly interested in the Rajshahi division, where the Bogra and Sirajganj districts are. Indeed, the average monthly household expenditure for Rajshahi falls well below the national average and the division is the second poorest in Bangladesh at Taka 9,254.

Table 4: Monthly household consumption expenditures by administrative division (2010)

Division	Consumption expenditure
Total (National)	11,003
Barishal Division	9,826
Chittagong Division	14,360
Dhaka Division	11,643
Khulna Division	9,304
Rajshahi Division	9,254
Rangpur Division	8,298
Sylhet Division	12,003

Source: BBS (2011)

Although the consumption estimates just reported are different from those used in our study,¹⁷ they provide a frame of reference within which to assess our findings.

We can therefore move onto the comparative analysis of mean consumption. The estimate of mean per capita household consumption (total, food, non-food) is compared by module; the interpretation is relative to the benchmark used. Our prime interest is to understand differences between each of the two short consumption modules and the benchmark. Table 5 shows the results of pair-wise t-statistics between the benchmark and each short consumption module.

¹⁷ The estimates just reported are at the household level, at current prices and include festivities and rent (see Section 4 and Annex 2 for further details).

Table 5 Consumption expenditure per capita (monthly Taka) by consumption module (categorized vs HIES; itemized vs HIES); pairwise t-tests

	(1)			(2)			(3)		Difference			
	<i>Categorized</i>			<i>Itemized</i>			<i>HIES - Bogra and Sirajganj districts (benchmark)</i>		<i>HIES-categorized</i>	%	<i>HIES - itemized</i>	%
FOOD CONSUMPTION	<i>mean</i>		<i>sd</i>	<i>mean</i>		<i>sd</i>	<i>mean</i>	<i>sd</i>				
Total food consumption	2095.8	***	1172.2	1892.7	***	1240.9	1459.0	604.0	-636.8	-30.4	-433.7	-22.9
NON-FOOD CONSUMPTION												
Non-food consumption	1200.3	***	1263.3	915.7		1417.7	760.1	534.1	-440.2	-36.7	-155.5	-17.0
TOTAL CONSUMPTION												
Total consumption	3296.1	***	2222.8	2808.3	***	2183.2	2219.1	1002.0	-1076.9	-32.7	-589.2	-21.0
<i>Sample size</i>	93			97			440					

Average monthly per capita consumption is Taka 3,296, 2,808 and 2,219 in the categorized, itemized and the restricted HIES samples respectively.

On one hand, consumption estimates from the *categorized module* consistently over-estimate benchmark means in terms of food, non-food and total consumption; the differences (30, 37 and 33% respectively) are, in most cases, statistically significant at the 1 per cent level. On the other hand, the *itemized module* is not statistically different from the benchmark in terms of non-food consumption; however, it also overestimates food (by 23%) and as a consequence total consumption (by 21%).¹⁸

These preliminary findings are not consistent with the literature which finds that short consumption modules tend to underestimate rather than overestimate consumption. One possible explanation for such a puzzling finding is that the restricted HIES benchmark sample is poorer than the pilot samples – as supported by the initial statistics shown in Section 5 – and that therefore their reported consumption is lower. Moreover, using HIES consumption data from 2010 as a benchmark implicitly assumes that no change in consumption has happened between 2010 and 2012 in Bangladesh apart from inflationary changes; however, we noticed that there has been a clear trend towards an increase in living standards that might well have continued into 2012.

Also, according to the literature (Beegle et al., 2010), the itemized module performs better than the categorized one (i.e. it underestimates less). Our preliminary findings indicate that the categorized module overestimates more than the itemized module¹⁹ (i.e. the itemized module overestimates less). As shown in Annex 3 Table A1, such finding can also be linked to some differences between the two short consumption modules samples noted in Section 5, namely the fact that the categorized sample appears to be richer than the itemized one.²⁰

In order to confirm the results obtained from descriptive statistics in the previous table, the natural logarithm of consumption (total and by main component) is regressed on dummy variables that indicate the module type (the category left out is the benchmark module). The coefficients can then be interpreted as percent deviation in mean value from the excluded category, namely the benchmark. This is the same methodology applied in Beegle *et al.* (2010) where a randomized survey experiment was used, so no further control variables were included in their regression.²¹ However, the Global MIC5 Pilot was not randomized and regressions with (Table 6) and without (Table A3 in Annex 5) further control variables are run and reported.

Regressions without control variables are still reported in Annex 5, as the comparison of these two tables provides an indication of to what extent controls are soaking up the initial consumption difference observed. Findings should be consistent with the t-tests run in the previous section.

¹⁸ Tests in Table 5 are based on the full categorized and itemized samples. However, we recomputed these also based on a sub-sample (i.e. dropping 2% or 5% of the top observations); although the percentage difference between the short and the module benchmark reduces, differences still remain statistically significant. Notwithstanding the possible influence of outliers we retained them in our analysis given that the samples are very small.

¹⁹ Annex 4 also reports statistical differences in mean consumption between the two short consumption modules. As can be noticed however, differences in means for total consumption as well as for food and non-food consumption are not statistically significant.

²⁰ Annex 6 also reports a more detailed comparison of means at the disaggregated level (moving from aggregates to sub-aggregates). Results at the disaggregated level are discussed more in sub-section 6.1.3, Table 9.

²¹ Indeed, as they explain: "The addition of controls may introduce bias into the experimental estimator (Freedman, 2008)", (Beegle *et al.*, 2010).

However, the dependent variables are logged; therefore, given the non-linear transformation of consumption (logged), percentage deviations from the mean might depart from those reported in Table 5.²²

In the case of the categorized module, total food consumption is statistically significantly higher (around 30.5 per cent) than mean consumption collected using the benchmark module; the itemized module, after scaling, results in 17.8 per cent higher food consumption. The impact on non-food consumption is 24.6 per cent in the collapsed module; however, no statistically significant difference in non-food consumption is observed for the subset module. Overall, total consumption is found to be overestimated in both short modules: by 13.7 per cent in the itemized module and by 30.5 per cent in the categorized module. Percentage deviations in overall consumption from the categorized module are more than double those obtained with the itemized module. Results of logged per capita consumption, using the reduced HIES sample (Bogra and Sirajganj) as a benchmark, with further control variables, are reported in Table 6.

Table 6 Regressions of per capita consumption expenditure logged, by total, food, and non-food component (with added control variables)

	Bogra and Sirajganj <u>FOOD</u> <i>Food consumption</i>	<u>NON-FOOD</u> <i>Non-food consumption (excluding ceremonies and rent)</i>	<u>TOTAL</u> <i>Total consumption (excluding ceremonies and rent)</i>
	(1)	(2)	(3)
Categorized	0.174*** (0.0520)	-0.0669 (0.0802)	0.105** (0.0522)
Itemized	0.154** (0.0597)	-0.216** (0.101)	0.0775 (0.0610)
Household size	-0.0867*** (0.0110)	-0.117*** (0.0146)	-0.0957*** (0.00893)
Age of household head	0.00122 (0.00126)	0.00718*** (0.00179)	0.00306*** (0.00118)
Electricity	0.0831** (0.0419)	0.150*** (0.0539)	0.110*** (0.0393)
Urban	0.0136 (0.0382)	-0.246*** (0.0530)	-0.0668* (0.0359)
Cattle	0.0308*** (0.0114)	0.0355** (0.0171)	0.0312*** (0.0110)
Goat	0.00913 (0.0124)	0.00508 (0.0190)	0.00545 (0.0118)
Chicken/duck	0.000132 (0.000156)	0.000269 (0.000209)	0.000118 (0.000137)
Radio	-0.00121 (0.0789)	0.156* (0.0937)	0.0593 (0.0711)
Television	0.157*** (0.0406)	0.0645 (0.0578)	0.119*** (0.0373)
Landline	-0.000838 (0.112)	-0.273*** (0.101)	-0.102 (0.0921)

²² Regressions were also run without the log transformation; in this case, the coefficients provide the actual difference between the benchmark and the module estimate (and therefore coincide with the 'difference column' in Table 5).

Refrigerator	0.208** (0.0939)	0.397*** (0.117)	0.295*** (0.0836)
Watch	0.104*** (0.0369)	0.146*** (0.0457)	0.119*** (0.0333)
Mobile	0.107*** (0.0396)	0.329*** (0.0499)	0.189*** (0.0366)
Bicycle	0.0253 (0.0357)	0.00607 (0.0524)	0.0156 (0.0337)
Motorcycle/scooter	0.0751 (0.0887)	0.357*** (0.135)	0.186** (0.0786)
Car/truck	-0.0140 (0.0988)	0.0330 (0.197)	-0.00923 (0.117)
Owns dwelling	0.0621 (0.0545)	0.0998 (0.0662)	0.0677 (0.0495)
Land for agriculture	-0.0172 (0.0358)	0.0503 (0.0515)	0.00496 (0.0332)
Animals	0.0377 (0.0407)	0.0858 (0.0554)	0.0583 (0.0376)
Tubewell	0.294*** (0.0924)	0.0716 (0.458)	0.152 (0.257)
Supply water	0.247 (0.154)	0.385 (0.498)	0.252 (0.279)
Islam	0.131 (0.0819)	-0.0441 (0.112)	0.0643 (0.0767)
Metal/tin roof	-0.184*** (0.0642)	-0.113 (0.0938)	-0.160*** (0.0552)
Household with children	-0.106** (0.0464)	-0.0376 (0.0622)	-0.0891** (0.0428)
Male head of the household	0.0814 (0.0551)	0.0105 (0.0973)	0.0442 (0.0518)
Education 1	-0.169** (0.0732)	-0.210 (0.203)	-0.187** (0.0867)
Education 2	-0.203*** (0.0779)	-0.357* (0.205)	-0.261*** (0.0900)
Education 3	-0.0769 (0.0739)	0.00343 (0.193)	-0.0528 (0.0838)
Constant	7.108*** (0.207)	6.800*** (0.504)	7.751*** (0.288)
Observations	629	629	629
R-squared	0.472	0.588	0.600

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The table shows that controlling for other factors, the impact of the module assigned on consumption is reduced (compared to table A3 in Annex 5). On the one hand, the categorized module is still found to overestimate total consumption although the percent deviation is now much lower (10.5 rather than 30.5). This result is mainly driven by an overestimation in food consumption (17.4% with respect to the benchmark); whereas the impact on non-food consumption is no longer statistically significant. On the other hand, the impact of the itemized module on total consumption is no longer statistically significant; however, disaggregating consumption it seems that this result is driven by biases in opposite directions concerning food and

non-food consumption. Indeed, total food consumption, after scaling, is statistically significantly higher than mean consumption collected using the benchmark module (15.4% higher), whereas non-food consumption is underestimated by 21.6 per cent.

6.1.2 Using predicted household consumption levels as a benchmark

As a further check in the validation process, this section of the paper predicts household consumption levels from the Bangladesh HIES 2010 to the two short consumption samples. Predicted consumption is then used as a benchmark. This methodology is a further step towards tackling the difficulties of comparing estimates across three non-equivalent samples.

Consumption regressions have been implemented in different contexts²³ and the estimated models are considered to have considerable predictive power. Basically, per capita consumption is predicted at the household level in our pilot survey using the information available on these households in the same dataset as well as the parameter estimates obtained from a consumption model using HIES 2010. Consumption is regressed on a set of variables that are considered to be correlated with consumption, household characteristics and assets that are comparable across surveys; coefficient estimates are then used as weights.

One of the weaknesses of this method is that it assumes that the relationship between consumption and its correlates remains stable over time (i.e. that the determinants of consumption do not change over time). Indeed, predicting consumption is expected to be more accurate when a fairly recent survey is available and when no major shock has occurred in the country. The assumption is 'riskier' the more dynamic the economy is and the greater the time-span between surveys.

Consumption regressions in the HIES sample (restricted to Bogra and Sirajganj) were run using as regressors variables common to both surveys. The dependent variable is logged. The regressions has fairly high R^2 (0.653).²⁴

Consumption is then predicted in each of the two short consumption modules using the coefficients of the regression as weights; once predicted, log consumption is transformed back into Taka.²⁵ It is therefore tested whether mean consumption in each of the short consumption modules statistically differs significantly from the one predicted from the consumption regression based on the HIES sample.²⁶ Tests are reported in Table 7.

²³ For instance, this approach has been used to link survey and census data for poverty mapping. "This approach of imputing consumption from one data source into another data source has been applied by Elbers, Lanjouw and Lanjouw (2002, 2003) in a number of countries in the context of producing "maps" of poverty and inequality by imputing consumption from a household survey into the population census" (Kijima and Lanjouw, 2003).

²⁴ Consumption regressions are reported in Table A6 in Annex 7.

²⁵ We use the parameters estimated ($\hat{\beta}$) in our consumption regression model to generate predictions of consumption per capita (\hat{c}_j) for every household j using the following formula: $\hat{c}_j = e^{(\hat{\beta} \cdot x_j + \hat{\sigma}^2/2)}$. Basically, in order to transform the log of consumption back to Taka, the term $\hat{\sigma}^2/2$, where $\hat{\sigma}$ is the estimated standard error of the regression, is added; this, as reported in Greene (1997), is required due to the lognormal transformation of the dependent variable (Simler et al., 2004).

²⁶ Our standard errors were recomputed to take into account the fact that consumption is predicted (see formulas in Simler et al., 2004:91).

Table 7. Actual and predicted consumption for the categorized and itemized modules (pair-wise t-tests)

		(1) Categorized			(2) Itemized			Difference		Difference	
		<i>Actual</i>	<i>Predicted</i>	<i>Pair-wise t-test</i>	<i>Actual</i>	<i>Predicted</i>	<i>Pair-wise t-test</i>	HIES-categorized	%	HIES-itemzed	%
Consumption (excluding ceremonies and rent)	<i>mean</i>	3296.1	3003.9		2808.3	2488.1		-292.2	-8.9	-320.2	-11.4
	<i>sd</i>	2222.8	2493.8		2183.2	1048.4					
<i>Sample</i>		<i>93</i>	<i>93</i>		<i>97</i>	<i>97</i>					

Predicting consumption allows controlling for some important determinants of the outcome variable that differ across the three samples; its application should therefore help to reduce the confounding factors. Findings now differ from what was observed in the previous sub-section (6.1.1) where estimates were being compared across non-equivalent samples. Indeed, although mean consumption in the categorized and itemized modules is still slightly higher than that predicted (i.e. the benchmark), differences in means are no longer statistically different.

So, why are findings different from those in the previous sub-section? In the first part of this paper we were comparing samples that differ in characteristics which are important in determining the level of consumption. The differences in estimates observed could therefore not be unequivocally explained by the different ability of short modules to capture consumption due to their module design and/or length; indeed, comparing three non-equivalent samples could have driven the differences observed. However, as we reduce the sampling bias by controlling for important characteristics that determine consumption and that vary across the subsamples, we are able to better isolate the impact of module design (attribution).

6.1.3 Propensity score matching approach

This section applies a non-experimental or quasi-experimental design technique. In particular, propensity score matching is used to try to pick an ideal comparison that matches the subsample to which each short consumption module was administered (i.e. the treatment group) using data from a larger survey (in this case the HIES 2010). HIES 2010 data is therefore again used to define the benchmark.

Hereafter, the main steps in the analysis are described. Basically, HIES 2010 (restricted to Bogra and Sirajganj²⁷) and the pilot dataset (either the itemized or the categorized) are pooled and a probit model of ‘participation’ is estimated as a function of all the variables in the data that are likely to determine participation.²⁸ Usually the variables to be included should be those that simultaneously influence the treatment status and the outcome variable; in this case variables that determine (per capita) consumption are used.

The predicted values of the probability of ‘participation’ from the probit regression are then computed (i.e. “propensity scores”) for each household sampled in the pilot (the short consumption module) and the HIES 2010.

For each individual in the pilot sample, the observation in the HIES 2010 sample that has the closest propensity score, as measured by the absolute difference in scores, is found. This is called the “nearest neighbour”²⁹. We can find the n nearest neighbours; in this case we used the three nearest neighbours. These observations will represent the benchmark.

With the *nearest neighbour (NN)* method, all treated units find a match. However, some of these matches may be fairly poor if the nearest neighbour for some treated units has a very different propensity score. *Caliper matching* is therefore used to tackle this problem; basically the maximum difference between a treated case’s propensity score and its matching control case (0.01 in this

²⁷ The sample is purposively restricted to allow matching only within the same geographic area.

²⁸ The dependent variable participation is equal to 1 if the household was administered the short consumption module and equal to 0 otherwise (i.e. if household was administered HIES 2010).

²⁹ The nearest neighbour is only one of the possible propensity score matching methods tried. Results are available upon request.

case) is set. A common support is then imposed by dropping treatment observations whose propensity scores are higher than the maximum or less than the minimum propensity score of the controls. Finally, matching is applied *with replacement*, meaning that the same control cases could be used more than once as matches to treated cases.

The *mean value of consumption* for the three nearest neighbours is then calculated. The difference between that mean and the actual value for the treated observation (from pilot) is the estimate of the under/over-estimation due to the short (versus long) consumption module. The mean of these individual under/over-estimations is calculated to obtain the average overall under/over-estimation. This could potentially be stratified by some variable of interest, such as education, location, etc. however the pilot samples are fairly small and limit the analysis.

The ‘balancing check’ was carried out before interpreting results. Basically, the balancing check tests whether the means of each covariate differ between the treated and the control group; there should no longer be any statistically significant differences (as can be seen in Annex 8, Tables A7 and A8 that report the characteristics of the matched groups from HIES and each pilot sample). The percentage bias is often used as a measure of balancing. In this paper, the mean percentage bias after matching is below 7-8% for most covariates.

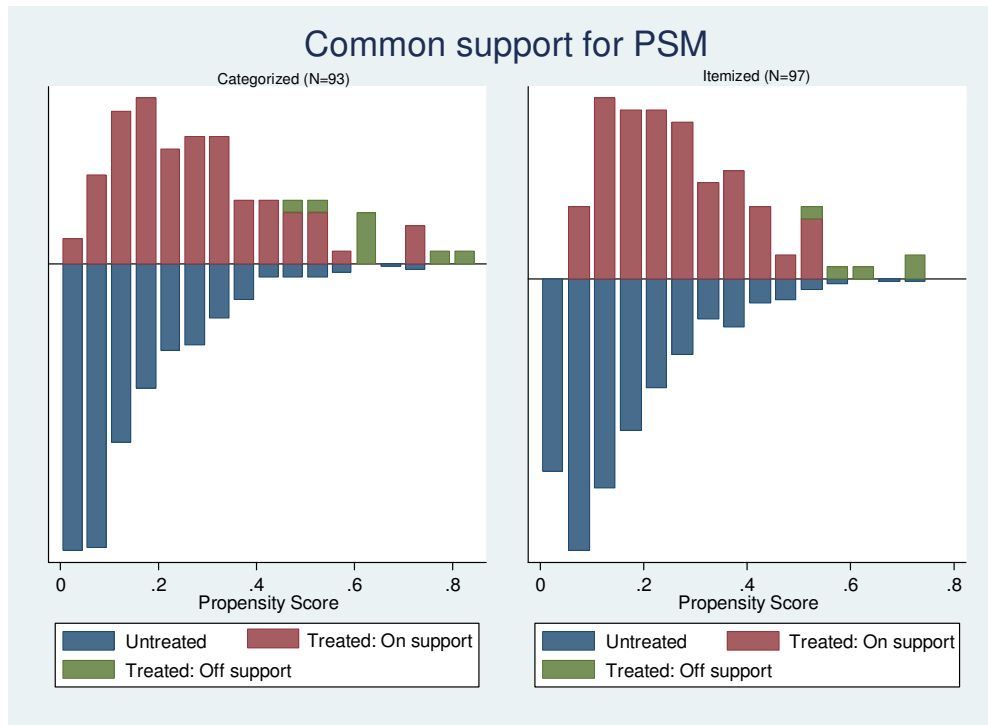
The table also provides a summary of the distribution of the absolute bias before and after matching. The average percentage absolute bias before matching was 18.7% and 19.3% for categorized and itemized respectively. After matching, it becomes 6.4% and 5.3%. Therefore matching reduces initial unbalancing (see Annex 8).

Figure 2 also shows separately the balancing scores for the categorized and for the itemized samples to assess graphically if the assumption of ‘common support’ holds. The graphs show the treated cases in red above and the control cases (HIES restricted sample) in blue below; the dropped treatment observations are in green.³⁰

Although control cases have more propensity scores towards the lower end of the distribution, there are treatment and control cases almost everywhere; we can therefore conclude that the assumption holds, as there is an overlap of the propensity scores of the treated and untreated (i.e. short consumption samples and HIES restricted sample).

³⁰ See Annex 8 for more information.

Figure 2. Common support for propensity score matching (categorized vs HIES restricted sample; itemized vs HIES restricted sample)



Results (average treatment effects) are reported in Table 8.³¹ The table shows, for the main consumption aggregates, the “ATT,” which is the average treatment effect for the treated cases. The “unmatched” results – namely the difference between the average consumption for the treated and the control group, before any matching is done (i.e. the difference in the raw data), is not reported.

³¹ It should be reminded that propensity score matching (PSM) controls for differences in observables, therefore selection bias due to unobserved differences between treated and control group may still occur.

Table 8. Average treatment effects by main consumption aggregates and consumption module

	(1) <i>Categorized</i>							(2) <i>Itemized</i>					
	Sample	Treated	Controls	Difference	S.E.	T-stat		Sample	Treated	Controls	Difference	S.E.	T-stat
TOTAL CONSUMPTION													
Consumption (excluding ceremonies and rent)	ATT	3267.35	2960.59	306.76	270.04	1.14		ATT	2977.41	2706.30	271.11	248.17	1.09
FOOD CONSUMPTION													
Total food consumption	ATT	2123.19	1742.54	380.65	147.65	2.58 **		ATT	2035.80	1729.85	305.95	155.53	1.97 *
Food consumption <i>within</i> the household	ATT	2017.24	1672.84	344.40	140.50	2.45 **		ATT	1850.42	1684.63	165.79	148.05	1.12
Food consumption <i>outside</i> the household	ATT	169.92	67.56	102.36	33.86	3.02 ***		ATT	230.48	43.83	186.64	43.47	4.29 ***
NON-FOOD CONSUMPTION													
Non-food consumption (excluding ceremonies and rent)	ATT	1144.16	1218.04	-73.88	144.92	-0.51		ATT	941.61	976.45	-34.84	126.19	-0.28

Findings show that once the ‘treated’ group (categorized or itemized) is matched to the control, many of the statistically significant differences that were noticed in the initial section disappear.

Differences in means for *aggregate* consumption are no longer statistically significant. This holds for both the itemized and the categorized modules. This finding is important as it points to the fact that the analysis carried out in the first part of the paper (sub-section 6.1.1) is biased by confounding factors that do not allow us to conclude that consumption in the itemized and categorized sample was higher than the benchmark due to differences in the performance of modules (or at least not only); this result was at least partially driven by differences in the characteristics of the samples (i.e. samples are non-equivalent).

Once consumption is disaggregated (food and non-food) however, there are still some significant differences worth noticing.

Focusing on *food* consumption, although the difference between the mean of the control group and that of the treated group has decreased with matching, some differences are still statistically significant. This is true for both the categorized and the itemized consumption modules. However, whereas both food consumption within and outside the household are overestimated in the categorized module, in the itemized module differences in mean food consumption within the household are not statistically different; food consumption outside the household however is heavily overestimated in the itemized module (almost six times).

Food consumption outside the household is collected through the same question in both short consumption modules and it is therefore not surprising to find consistent results. In general, it is difficult for any survey respondent to appropriately report on individual household members’ consumption that occurs outside of its purview; in our short consumption modules this might have been exacerbated by the fact that the respondent is not necessarily the most knowledgeable adult member in terms of consumption but more generally any knowledgeable adult (more than 15 years old) household member.

Table 9 reports findings for several consumption sub-aggregates: results are reported by specific food category for the categorized module and by food item for the itemized module. It is interesting to see that most food categories/items that statistically significantly differ from the benchmark do so by over-estimating the ‘true’ value of consumption. One of the reasons that could be put forward to explain certain discrepancies in food sub-aggregates is seasonality; just to provide an example, it was mango season at the time when the pilot was carried out and therefore mango consumption estimates would be expected – as in fact happens – to be higher than in the annual HIES.

Moreover, our short consumption modules stressed the importance of gifts and assistance in computing consumption. A separate question about gifts/assistance was asked for each individual food item/category; in our sample poor families also report meat consumption due to the fact that even the most disadvantaged households receive meat as assistance/gifts during one of the religious festivals in Bangladesh; it is difficult to know whether this holds in the benchmark where one question is asked for consumption overall (including purchases, wages in kind, self-production and gifts).

Among food sub-aggregates, however, both the 'cereals' (in the categorized module) and the 'rice and rice flour' (in the itemized module) categories are underestimated rather than over-estimated. This particular result could be driven either by the fact that in both modules questions about these items/categories come first – at the stage in which respondents still need to fully understand how the questionnaire functions – or simply by the fact that these are the major food items/categories consumed and as such are more likely to be forgotten (recall) and therefore under-estimated.

The HIES 2010 report indicates however that between 2005 and 2010 the share of food consumption relating to cereals has been decreasing; this trend could have continued into 2012.

As far as *non-food* consumption is concerned, at the aggregate level differences are no longer statistically different for either the categorized or the itemized module, whatever the definition used.

Disaggregating (Table 9) non-food consumption further, it can be seen that most of the sub-aggregates are still statistically significantly different for the itemized module (health and transport are overestimated whereas housing and other non-food expenditures are underestimated); however, after matching, education seems to be consistent with the results from the benchmark.

Table 9. Average treatment effects by consumption sub-aggregates and consumption module (ATT)

	(1) Categorized						(2) Itemized					
	Treated	Controls	Difference	S.E.	T-stat		Variable	Treated	Controls	Difference	S.E.	T-stat
FOOD CONSUMPTION												
Cereals and cereal products	461.26	618.80	-157.54	29.43	-5.35	***	Rice and rice flour	461.41	593.98	-132.58	54.05	-2.45 **
Roots and tubers	134.95	128.53	6.42	10.74	0.6		Soyabean oil	90.09	57.10	32.99	9.05	3.64 ***
Legumes/pulses and their products	48.60	41.03	7.57	5.84	1.3		Milk	76.78	78.22	-1.44	13.72	-0.11
Nuts and seeds	9.32	0.00	9.32	2.19	4.25	***	Potato	62.10	45.74	16.35	5.59	2.93 ***
Vegetables and vegetable prod	171.02	73.80	97.22	14.06	6.92	***	Chilies	18.25	31.64	-13.40	2.15	-6.23 ***
Fruits	206.62	79.74	126.88	32.39	3.92	***	Masoor (Lentil)	36.60	28.18	8.42	4.45	1.89 *
Sugars, Sweets and Syrup	79.63	44.44	35.18	17.00	2.07	**	Eggs	48.01	36.10	11.91	10.71	1.11
Meats, Poultry and their prod	275.74	198.31	77.43	45.16	1.71		Wheat and wheat flour	16.67	26.73	-10.06	5.64	-1.78
Eggs	58.26	32.34	25.92	9.35	2.77	***	Onion	28.12	27.58	0.54	2.80	0.19
Fish and Shellfish	243.37	206.33	37.03	26.69	1.39		Turmaric	17.29	17.38	-0.09	1.76	-0.05
Milk and Milk products	104.89	84.97	19.91	16.01	1.24		Mango	78.32	20.24	58.08	15.56	3.73 ***
Oils and fats	115.25	80.25	35.00	9.62	3.64	***	Beef	159.84	95.30	64.54	26.00	2.48 **
Beverages	34.06	8.51	25.55	8.63	2.96	***	Chick ducks and birds	80.09	59.30	20.80	19.83	1.05
Others	74.27	75.78	-1.51	6.33	-0.24							
NON-FOOD CONSUMPTION												
Education	231.36	263.94	-32.58	53.48	-0.61		Education	193.76	142.25	51.51	49.51	1.04
Health	193.42	104.14	89.28	51.73	1.73	*	Health	193.87	79.24	114.64	52.08	2.2 **
Clothing and footwear	262.31	168.80	93.51	40.40	2.31	**	Housing <i>excluding rent</i>	220.96	328.66	-107.70	39.79	-2.71 **
Housing <i>excluding rent</i>	145.45	281.58	-136.13	28.12	-4.84	***	Transport & Leisure and Culture related activities	229.50	107.68	121.82	39.65	3.07 ***
Transport	188.22	157.22	31.00	30.64	1.01		Other non-food expenditures	223.06	331.76	-108.70	38.01	-2.86 **
Communications	125.74	80.30	45.44	18.70	2.43	**						
Leisure and culture related activities	36.97	2.82	34.15	5.47	6.24	***						
Cigarettes, chewgoods and other tobacco products	114.68	101.61	13.06	23.95	0.55							
Personal care and services	82.17	63.00	19.17	10.26	1.87	*						
Kitchen and toilet products	33.18	21.90	11.28	3.81	2.96	***						

Turning to the categorized module, health is overestimated, but the remaining annual consumption expenditure, education, is not statistically significantly different from the control group. It seems that monthly consumption expenditures are more difficult to gather correctly; indeed although clothing, transport and cigarettes are similar to the control group, communications, leisure, personal care and 'kitchen and toilet products' are overestimated, while housing expenditures are underestimated. In part, some of these differences could be driven by the fact that the components of certain categories are classified by the respondent under a different category; thus they might still be picked up but misclassified. For instance, during the fieldwork it often happened that housing bills would be included with rent. Some non-food consumption categories that occur outside the purview of the respondent (such as communications) are again problematic as they might have been 'guesstimated'. Finally, and applying to both the categorized and itemized modules, although it was stressed that health should not have included inpatient expenditures, it might have been difficult for respondents to systematically distinguish these expenditures.

6.1.4 Consumption levels: summary table of findings

Results obtained using different approaches are reported in Table 10.

Table 10. Consumption levels: summary table of findings

	% difference [categorized vs HIES restricted]				% difference [itemized vs HIES restricted]		
	<i>Consumption</i>				<i>Consumption</i>		
	Total	Food	Non-food		Total	Food	Non-food
Raw/non-equivalent	32.7%***	30.4%***	36.7%***	Raw/non-equivalent	21%***	22.9%***	17.0%
Regression w. controls	10.5%**	17.4%***	-6.7%	Regression w. controls	7.8%	15.4%**	-21.6%**
Predicted consumption	8.9%	-	-	Predicted consumption	11.4%	-	-
PSM	9.38%	17.93%**	-6.46%	PSM	9.10%	15.03%*	-3.70%

The table shows that, as expected, the differences in consumption estimates observed between the benchmark and each short consumption module is reduced as we apply different methodologies that control for the use of non-equivalent samples

Total consumption is around 9-10% higher for the categorized module and 8-11% higher for the itemized module. Using propensity score matching, total consumption is not statistically significantly different for the categorized and itemized modules. Basically this means that once we eliminate existing confounding factors, we find that the impact of module assignment is not statistically significantly different, namely consumption estimates from each short consumption module are not statistically significantly different from the benchmark used each time.

If, however, we disaggregate total consumption, we notice some statistically significant differences in food consumption; indeed, the categorized and itemized modules overestimate food consumption by 18% and 15% respectively. In the itemized module the overestimation is driven by food consumption outside the household (indeed food consumption within the household is not

statistically significantly different), whereas in the categorized module both food consumption within and outside the household are overestimated. Non-food expenditures tend to be underestimated.

6.2 Poverty Estimate

This *second part* of the data analysis (Section 6.2) compares poverty estimates. Similarly to what was done in section 6.1 for consumption levels, poverty estimates from each short consumption module are assessed against a benchmark.

First poverty estimates are compared (sub-section 6.2.1) to the benchmark defined as the reduced HIES sample covering Bogra and Sirajganj (n=440) (i.e. raw comparison across non-equivalent samples). Poverty estimates are *then* computed on the basis of household total consumption as predicted in sub-section 6.1.2. Actual poverty rates are compared to the predicted ones that are used as a benchmark. Finally, in the *last part* of the analysis (Section 6.2.3), propensity score matching is applied. The section concludes again with a summary table of results.

6.2.1 Preliminary comparison of estimates across the three samples: differences in poverty rates (poverty analysis)

This section explores whether the poverty headcount (absolute³²) computed from each of the two piloted short consumption modules is statistically significantly different from the benchmark (the restricted HIES sample).

Before starting the comparative analysis, we provide a short overview of poverty levels in Bangladesh according to the Bangladesh HIES Report. Table 11 reports the incidence of poverty at national, rural and urban levels for 2010 and 2005 using both an upper and lower poverty line.

Table 11. Poverty rates in Bangladesh, 2005 and 2010, using lower and upper poverty line

	Upper poverty line			Lower Poverty line		
	National	Rural	Urban	National	Rural	Urban
2010	31.5	35.2	21.3	17.6	21.1	7.7
2005	40	43.8	28.4	25.1	28.6	14.6

The higher consumption observed in 2010 with respect to 2005 (as discussed in sub-section 6.1.1), translated into lower absolute poverty rates. In 2010, using the upper poverty line, headcount rates were estimated at 31.5, 35.2 and 21.3 per cent at the national, rural and urban levels respectively. In 2005, these rates were much higher; indeed, the incidence of poverty decreased by 8.5, 8.6 and 7.1 percentage points at the national, rural and urban level respectively during the period 2005 to 2010. A similar pattern is noticeable when using the lower poverty line. Headcount rates clearly vary broadly by administrative division. We are particularly interested in Rajshahi New that had in 2010 an incidence of poverty of 16.8 at the divisional level and of 17.7 and 13.2 at the

³² Relative poverty rates were also computed (using half and 2/3 of the median and half of the mean) although results are not reported in this paper (it is common practice to use absolute poverty in developing countries where the incidence of poverty is high).

rural and urban (divisional) levels respectively using the lower poverty line. Poverty rates in Rajshahi New for the same year but using the upper poverty line are 29.8, 30.0 and 29.0 at the divisional, rural and urban level respectively.

Let's now move on to the comparison of poverty estimates in our samples. As levels of consumption estimated in sub-section 6.1.1 were found to be over-estimated in the previous section, we expect absolute poverty rates to be under-estimated. Table 12 shows FGT measures (headcount, poverty gap and squared poverty gap) and pairwise t-tests.

Table 12. FGT measures by consumption module (pairwise test: categorized vs HIES, itemized vs HIES)

CONSUMPTION (excluding festivities and rent)			
	(1)	(2)	(3)
	<i>Categorized</i>	<i>Itemized</i>	<i>HIES - Bogra and Sirajganj</i>
FGT0 - Poverty headcount			
Lower poverty line	17.03 (3.75)	20.45 (4.55)	24.25 (9.99)
Upper poverty line	26.76 (5.21)	** 35.23 (6.59)	40.93 (15.23)
FGT1 - Poverty gap			
Lower poverty line	3.25% 0.09	4.23% 0.11	4.59% 0.10
Upper poverty line	6.26% 0.13	7.99% 0.15	9.14% 0.15
FGT2 - Squared poverty gap			
Lower poverty line	0.92% 0.03	1.43% 0.06	1.29% 0.04
Upper poverty line	2.12% 0.06	2.87% 0.08	3.00% 0.06
<i>Sample</i>	93	97	440

We first focus on the *incidence* of poverty. *Absolute poverty headcount rates* should be lower when using a lower poverty line and higher when using a less stringent poverty line (upper poverty line). This pattern can be noticed not only for the HIES but also for the two short consumption modules. *Although absolute poverty rates* tend to be slightly lower – as expected - with respect to the benchmark, differences are never statistically significant for the itemized module notwithstanding whether the lower or upper poverty line is applied. As for the categorized module, absolute

poverty rates are not significantly different from the benchmark when using the lower poverty line; however, absolute poverty estimates are lower when applying the upper poverty line.

The table also reports the results for the poverty and squared poverty gap. In both pilot samples, these measures are not statistically different from the ones computed for the benchmark; basically, the use of a short consumption module does not make a difference (statistically significant) in measuring the depth and severity of poverty.³³

6.2.2 Using predicted household consumption levels to compute poverty rates

In sub-section 6.1.2 we estimated a consumption correlates model and predicted per capita consumption at the household level. We now compute the (predicted) poverty status of each household by comparing the predicted level of consumption to the poverty line.³⁴

Table 13. FGT measures by consumption module (pairwise test: categorized vs HIES, itemized vs HIES). Actual versus predicted

	<i>(1) Categorized</i>		<i>(2) Itemized</i>		<i>(3) HIES - Bogra and Sirajganj</i>	
	<i>Actual</i>	<i>Predicted</i>	<i>Actual</i>	<i>Predicted</i>	<i>Actual</i>	<i>Predicted</i>
FGT0 - Poverty headcount						
Lower poverty line	17.03 (3.75)	16.1 24.0	20.45 (4.55)	20.2 25.1	24.25 (9.99)	24.5 27.4
Upper poverty line	26.76 (5.21)	28.1 31.8	35.23 (6.59)	34.6 32.4	40.93 (15.23)	40.4 33.0
FGT1 - Poverty gap						
Lower poverty line	3.25% 0.09	2.94 5.23	4.23% 0.11	3.72 5.89	4.59% 0.10	4.82 7.26
Upper poverty line	6.26% 0.13	6.20 9.05	7.99% 0.15	7.75 9.68	9.14% 0.15	9.51 10.90
FGT2 - Squared poverty gap						
Lower poverty line	0.92% 0.03	0.80 1.58	1.43% 0.06	1.03 1.90	1.29% 0.04	1.42 2.69
Upper poverty line	2.12% 0.06	1.97 3.32	2.87% 0.08	2.49 3.75	3.00% 0.06	3.19 4.63
<i>Sample</i>	<i>93</i>	<i>93</i>	<i>97</i>	<i>97</i>	<i>440</i>	<i>440</i>

Predicted poverty in each of the short consumption modules is then used as a benchmark to which compare actual estimates computed from the pilot samples data³⁵. We find no statistically significant difference once we compare actual data to the respective benchmarks. Indeed, poverty

³³ Stochastic dominance analysis was carried out and is available upon request.

³⁴ Indeed, the “poor” dummy was computed based on predicted consumption. We therefore had to take into account the fact that consumption is predicted and therefore comes with a forecast error. In order to compute poverty rates based on predicted consumption we use the formulas as for Simler et al. 2004 (see footnotes 26 and 27).

³⁵ Actual and predicted values are however reported also for the HIES sample.

rates obtained using consumption data collected through short consumption modules are very similar to the ones computed using consumption predicted from HIES on the basis of a number of household characteristics and assets.

The difference in the incidence of poverty is around a percentage point for the categorized module (predicted poverty is 16% against 17% from actual data when using the lower poverty line; predicted poverty is 28% against 27% from actual data when using the upper poverty line) and for the itemized module (predicted and actual poverty is 20% when using the lower poverty line and 35% when using the upper poverty line).

This table therefore seems to highlight that short consumption modules perform at least as well as consumption prediction models when it comes to poverty incidence, depth and severity estimates.

6.2.3 Propensity score matching approach

In sub-section 6.1.3 we discussed differences in consumption levels between the two short consumption modules and the benchmark, using the propensity score matching approach; we now turn to comparing poverty measures based on the same methodology. Table 14 reports average treatment effects for absolute poverty measures.

Table 14. Absolute poverty rates - Average treatment effects by main consumption aggregates and consumption module

<u>CATEGORIZED</u>						
	Sample	Treated	Controls	Difference	S.E.	T-stat
FGT0 - Poverty headcount						
Lower poverty line	ATT	0.1529	0.1549	-0.0020	0.0525	-0.04
Upper poverty line	ATT	0.2588	0.2176	0.0412	0.0632	0.65
FGT1 - Poverty gap						
Lower poverty line	ATT	0.0242	0.0215	0.0028	0.0107	0.26
Upper poverty line	ATT	0.0533	0.0500	0.0033	0.0167	0.2
FGT2 - Squared poverty gap						
Lower poverty line	ATT	0.0061	0.0052	0.0009	0.0036	0.24
Upper poverty line	ATT	0.0162	0.0148	0.0014	0.0065	0.21
<u>ITEMIZED</u>						
	Sample	Treated	Controls	Difference	S.E.	T-stat
FGT0 - Poverty headcount						
Lower poverty line	ATT	0.1957	0.1359	0.0598	0.0523	1.14
Upper poverty line	ATT	0.3370	0.2518	0.0851	0.0626	1.36
FGT1 - Poverty gap						
Lower poverty line	ATT	0.0407	0.0193	0.0214	0.0130	1.64
Upper poverty line	ATT	0.0773	0.0490	0.0283	0.0184	1.54
FGT2 - Squared poverty gap						
Lower poverty line	ATT	0.0134	0.0044	0.0091	0.0059	1.53
Upper poverty line	ATT	0.0275	0.0135	0.0140	0.0086	1.62

Table 14 shows that after matching there are no more significant differences between the treated and control group in any of the FGT absolute poverty measures. This holds for both the categorized and the itemized modules, when applying either a lower or an upper poverty line.

6.2.4 Poverty estimates: summary table of findings

Table 15 attempts to sum up findings on poverty estimates. As can be noted (absolute) poverty estimates obtained through the categorized and itemized short consumption modules do not differ significantly from the benchmarks used notwithstanding the methodology used. The only discrepancy concerns the headcount rate for the categorized module when using the upper poverty line. However, as already mentioned the raw comparison does not take into account that samples are non-equivalent.

Table 15 - Discrepancies (in percentage points) between pilot poverty estimates and the benchmark by methodology

	Categorized			Itemized		
	Raw/non-equivalent	Predicted consumption	PSM	Raw/non-equivalent	Predicted consumption	PSM
FGT0 - Headcount						
Lower poverty line	-7.22	0.90	0.20	-3.80	0.27	5.98
Upper poverty line	-14.17 **	-1.30	-4.12	-5.70	0.63	8.51
FGT1 - Poverty gap						
Lower poverty line	-0.01	0.30	-0.28	0.00	0.51	2.14
Upper poverty line	-0.03	0.10	-0.33	-0.01	0.24	2.83
FGT2 - Squared poverty gap						
Lower poverty line	0.00	0.10	-0.09	0.00	0.40	0.91
Upper poverty line	-0.01	0.20	-0.14	0.00	0.38	1.40

6.3 Distributional Estimates

The previous sub-sections (6.1 and 6.2) have focused on mean differences in consumption and poverty. But it is also important to investigate differences in distributional measures and/or in how households are ranked according to the different instruments used. This section aims to analyse these differences.

First a raw comparison of consumption deciles between the short pilot modules and the HIES restricted sample is provided (sub-section 6.3.1); this allows us to see whether what was observed for mean consumption applies also across the distribution. Such analysis is preliminary, due not only to the fact that we are comparing non-equivalent samples but also because the samples are very small.

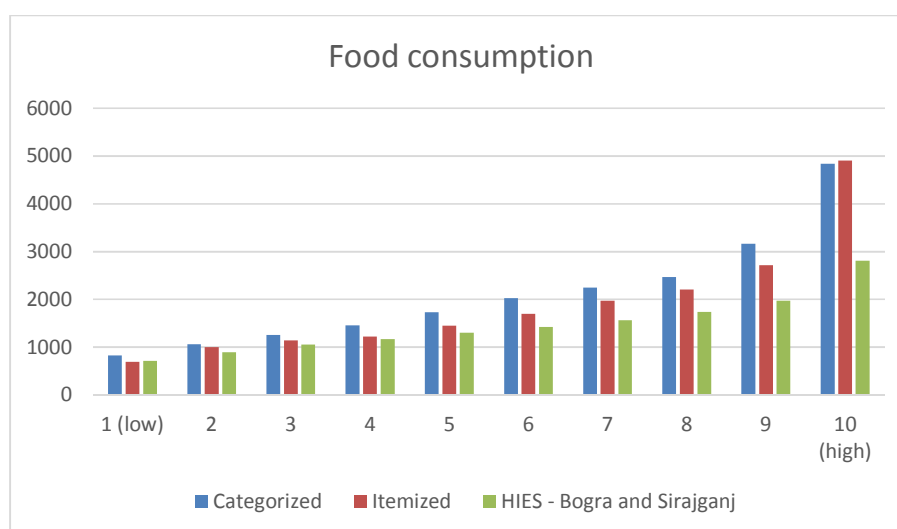
We therefore attempt to see whether consumption from the pilot modules provides the same classification of households by quintile. Indeed, even if there were a bias, if this were spread uniformly or linearly, it would not influence the ranking. In each pilot sample, we thus compute the percentage of households that fall within the same quintile according to the actual level of

consumption as collected with the 2 short modules and the predicted level of consumption (sub-section 6.3.2) or the widely used wealth index (sub-section 6.3.3).³⁶ Finally, a summary table of results is provided (sub-section 6.3.4).

6.3.1 Preliminary comparison of estimates across the three samples: differences in distributional estimates (inequality analysis)

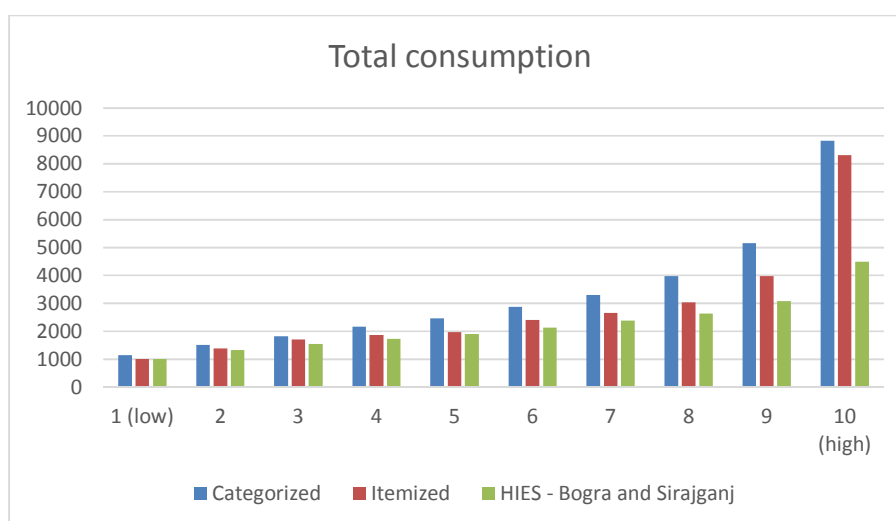
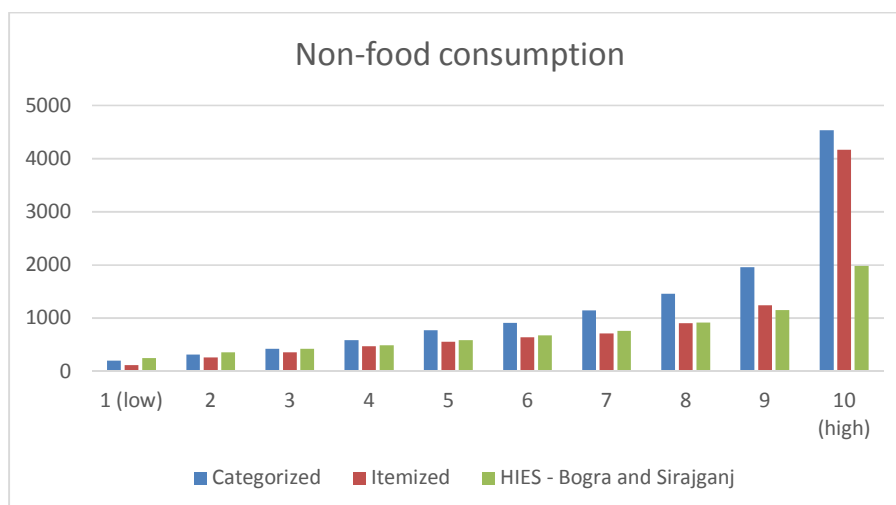
In this section, a raw comparison of consumption deciles between the short and long consumption modules is provided.³⁷ Figure 3 report deciles for food, non-food and aggregate consumption.

Figure 3. Food, non-food and total consumption deciles by consumption module



³⁶ It was not possible to use PSM to investigate distributional measures and/or ranking. Indeed, using this methodology, short consumption modules estimates are compared to an artificial sample with an artificial distribution. As we apply matching with replacement, one observation can be used more than once in the control group possibly biasing our estimates; moreover, for the control group, the mean of three observations (three nearest neighbours) is computed reducing somehow further the power of outliers and possibly flattening the control distribution; finally, as off support observations are dropped, further bias is introduced in the estimates derived from the short consumption modules.

³⁷ In Annex 9, in order to see whether distributions have similar shapes and/or are similarly skewed, consumption kernel density functions are visually compared; finally, Gini coefficients and other inequality measures are compared.



The mean of each decile of the categorized module tends to be higher than the benchmark. This holds true for food, non-food and total consumption; the only exceptions being the bottom deciles in the non-food consumption that are found to be lower than in the HIES restricted sample. The absolute difference between the HIES and the categorized modules is increasing in deciles; the same pattern, with few exceptions, is observed for the relative differences; there therefore seems to be a correlation between consumption levels and the divergence in consumption estimates (i.e. the overestimation within wealthier groups is found to be higher). It should again be stressed that such analysis is preliminary not only due to the fact that we are comparing non-equivalent samples but also because samples are very small.

6.3.2 Ranking based on predicted versus actual consumption

In this sub-section the ranking of the households is analysed. Basically, we show for each sample how the ranking provided by actual and predicted consumption differs; this is done in Table 16 which reports the cross tabulations between actual and predicted (i.e. benchmark) consumption

for each of the two short consumption modules and the reduced HIES sample (Bogra and Sirajganj districts).

Table 16. Cross-tabulation of actual and predicted (benchmark) consumption by module type

Consumption (excluding ceremonies and rent)									
(1)									
Categorized									
	Poorest	Second	Middle	Fourth	Richest	Total	hhs	%	
Poorest	8	6	5	1	0	19			
Second	5	4	6	2	1	18			
Middle	4	4	5	6	1	19			
Fourth	3	2	4	8	2	19			
Richest	0	3	0	2	13	18			
Total	20	18	19	19	18	93			
Falling in the same quintile							38	41%	
Falling within +/- 1 quintile							71	76%	
Sample						93			
Spearman's Rho 0.5801***									
(2)									
Itemized									
	Poorest	Second	Middle	Fourth	Richest	Total	hhs	%	
Poorest	10	1	4	4	1	20			
Second	5	8	4	3	1	20			
Middle	1	3	6	6	2	18			
Fourth	3	4	4	5	5	20			
Richest	1	3	2	2	11	19			
Total	20	19	19	19	19	97			
Falling in the same quintile							39	40%	
Falling within +/- 1 quintile							68	70%	
						97			
Spearman's Rho 0.4132***									
(3)									
HIES (Bogra and Sirajganj)									
	Poorest	Second	Middle	Fourth	Richest	Missing	Total	hhs	%
Poorest	52	24	11	2	0	1	88		
Second	30	29	22	4	3	0	88		
Middle	7	22	32	23	3	0	88		
Fourth	0	11	16	33	27	0	88		
Richest	0	1	6	26	54	0	88		
Total	89	88	88	88	88	1	440		
Falling in the same quintile								201	46%
Falling within +/- 1 quintile								368	84%
							440		
Spearman's Rho 0.7539***									

Using the benchmark, around 46% of households are classified in the same quintile by both predicted and actual consumption. The percentage of households correctly categorized is slightly

lower for the two short consumption modules, 40% of households in the categorized sample and 41% of households in the itemized sample fall in the same quintile when predicted and actual consumption are used. Predicted and actual consumption therefore provide a similar classification of households by quintile across the three samples. When households that fall within +/- 1 quintile are considered, then the percentage of households correctly categorized is higher; 84% for the HIES benchmark and 76% and 70% respectively for categorized and itemized.

At the bottom of Table 16 Spearman's correlations between predicted and actual consumption for each of the samples are reported. In each case the coefficients are statistically significant. The correlation between the predicted and actual consumption is 0.6 and 0.4 for the categorized and itemized modules respectively. The correlation is 0.7 for the benchmark (HIES restricted); therefore fairly similar to the categorized sample but relatively higher than for the itemized one. Predicted and actual consumption seem closely related in the HIES restricted and in the categorized sample but less so in the itemized one.

6.3.3 Ranking based on wealth index against consumption

In this sub-section we analyse further the ranking of households by investigating in each sample whether and how the ranking provided by consumption aggregates from different modules differ from the one derived from the widely used wealth index. Although consumption and wealth index measure different concepts and therefore no perfect correlation is expected, low associations could still raise questions about validity. Such comparison is particularly interesting given that wealth indices are widely used in multi-topic surveys such as MICS and/or DHS.

First, for each of the three samples an asset index is computed separately. The wealth index is computed using factor analysis.³⁸ Through this statistical technique it is possible to assign weights to the indicator variables used. Included in the wealth index are: quality of drinking water source, of sanitation facilities, of floors, walls and roofing material and of fuel used for cooking; a set of household assets such as radio, television, refrigerator, as well as some household members' assets such as bicycle, mobile phone and so on; and the number of persons per sleeping room in the household. The variables included in the wealth index vary slightly across samples and different options have been tried in order to come up with a high Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy that captures the appropriateness of factor analysis. As close an overlap as possible was sought in the list of household characteristics and assets chosen across the three samples.³⁹

Pair-wise correlations between wealth score and consumption for each module have been computed in order to check how closely related the wealth index is with consumption. In the literature, wealth indices have often been found to be weakly related to consumption with correlation coefficients in the range of 0.2-0.4 (World Bank, nd). All correlations are statistically significant; the correlation between the wealth score and overall consumption is 0.62 for the benchmark and 0.66 and 0.52 for the categorized and itemized modules respectively.

The wealth index is then used to group households by wealth quintile. Table 17 cross-classifies, for each module, households by quintile based on per capita consumption and the wealth index. Using

³⁸ Principal component analysis was also used as an alternative statistical technique, however results did not differ.

³⁹ However, HIES and MICS questionnaires do differ.

the benchmark, around 37% of households are correctly categorized by quintile by both consumption and asset. The percentage of households that are classified in the same quintile is therefore relatively low; this is expected and simply points to the fact that “wealth is not a straight proxy for per-member expenditures” (Rutstein and Johnson, 2004).

Assessing the validity of the wealth index is, however, beyond the aim of this paper; here the interest is simply in seeing whether the wealth score and consumption provide a similar classification of households by quintile across the three samples. The percentage of households correctly categorized is slightly lower for the categorized (with 30%) and itemized (with 29%) modules. If we consider households that fall within +/- 1 quintile, then the percentage of households correctly categorized is higher; 78% for the HIES benchmark and 72% and 60% respectively for the categorized and itemized.

Table 17. Cross-tabulation of wealth index and consumption by module type

(1)							
Categorized							
	Poorest	Second	Middle	Fourth	Richest	Total	
Poorest	8	5	4	2	0	19	
Second	6	1	8	3	1	18	
Middle	5	4	4	5	1	19	
Fourth	0	6	2	4	6	18	
Richest	0	3	1	5	11	19	
Total	19	18	19	19	18	93	
Falling in the same quintile						28	30.0%
Falling within +/- 1 quintile						67	72.4%
Spearman's Rho	0.5714***						
Pairwise correlation	0.6588***						

(2)							
Itemized							
	Poorest	Second	Middle	Fourth	Richest	Total	
Poorest	6	5	3	3	4	20	
Second	4	5	2	4	5	19	
Middle	9	4	4	2	0	19	
Fourth	1	5	6	6	2	19	
Richest	0	1	5	5	9	19	
Total	20	20	18	20	19	97	
Falling in the same quintile						28	29.1%
Falling within +/- 1 quintile						58	59.8%
Spearman's Rho	0.2936***						
Pairwise correlation	0.5199***						

(3)

HIES - Rajshahi division						
	Poorest	Second	Middle	Fourth	Richest	Total
Poorest	42	26	14	4	2	89
Second	21	32	16	13	7	88
Middle	16	22	20	18	12	87
Fourth	6	7	29	25	22	88
Richest	3	1	9	29	44	88
Total	88	88	88	88	88	440
Falling in the same quintile						164 37.2%
Falling within +/- 1 quintile						345 78.4%
Spearman's Rho	0.5461***					
Pairwise correlation	0.6193***					

6.3.4 Ranking households: summary table of findings

Table 18 provides a summary of the findings from sub-section 6.3. In particular, the welfare measure obtained from our short consumption modules would rank households similarly to that obtained by predicted consumption and the wealth index. Table 18 reports, for each short consumption module, the percentage of households that fall within the same quintile according to the different welfare measures used. To provide a benchmark to compare these values with, we also report on the same percentages for the benchmark sample (restricted HIES).

Table 18. Ranking Households: summary table of discrepancies (in percentage points) between pilot poverty estimates and the benchmark by methodology

% of hhs correctly classified	[categorized]		[itemized]		[HIES - Bogra and Sirajganj]	
	vs predicted consumption	vs wealth index	vs predicted consumption	vs wealth index	vs predicted consumption	vs wealth index
in the same quintile	41%	30%	40%	29%	46%	37%
within +/- 1 quintile	76%	72%	70%	60%	84%	78%

The categorized and itemized consumption aggregates classify around 40% of households exactly in the same quintile as predicted consumption would; this percentage is not high but is fairly similar to what would be obtained in the benchmark sample by cross tabulating actual and predicted consumption by quintile (i.e. 46%).

If we compare the classification of households based on consumption versus the widely used wealth index we find that in the benchmark sample around 37% of households are categorized in the same quintile according to these two measures; this percentage is relatively low but not uncommon in the literature as the wealth index and consumption capture different concepts.

The categorized and the itemized module have similar, although slightly lower, percentages (at 30% and 29% respectively). Percentages are higher when households that fall within +/- 1 quintile are considered; this holds true for cross-tabulations either with predicted consumption or the wealth index.

7. ROBUSTNESS CHECKS AND LIMITATIONS

In order to check our results, three different definitions of non-food consumption and therefore total consumption have been used; as the analysis is consistent, we report on only one definition of consumption. Similarly, in the underlying analysis to this paper two HIES sub-samples were used as *benchmarks*: 1) the sub-sample of Bogra and Sirajganj districts (n=440) and 2) the sub-sample of the Rajshahi division (n=1580); for clarity of exposition, however, results are reported only for the Bogra and Sirajganj sub-sample.

Further analysis and robustness checks could include re-estimating and comparing estimates using an adult equivalent scale rather than per capita figures. In addition, consumption in HIES (benchmark) could be re-computed taking into account when exactly each household had been interviewed so as to use different and more specific time deflators.

Finally, there are a number of limitations in this study. No perfectly comparable benchmark was administered during the pilot. Ideally survey design should have been the same across the short and long consumption modules: the method of data capture, the level of respondent, as well as the length of the reference period, should have been consistent in order to isolate the impact of the length or level of detail in consumption questionnaires on the accuracy of the resulting estimates.⁴⁰ The HIES 2010 provides our second-best proxy to household consumption. However, several issues should be kept in mind when comparing our estimates to HIES 2010: first, using consumption data from 2010 implicitly assumes that no change in consumption has happened between 2010 and 2012 in Bangladesh apart from inflationary changes; second, we are comparing seasonal consumption estimates (from our sample) to the annual ones of HIES 2010; finally, samples are fairly small (around 100) and do not allow for useful disaggregation.⁴¹

8. CONCLUSIONS

The aim of the paper was to understand how short consumption modules fared relative to a longer and more detailed consumption module in terms of the accuracy of the resulting estimates. The objective is particularly challenging as the use of non-equivalent samples makes it difficult to assess the accuracy and reliability of the estimates obtained. In order to overcome the problem of comparability and reduce the bias that follows from the use of non-equivalent samples, the paper applies several methodologies.

Results are promising. Once we control for differences in sample characteristics, the short consumption modules line up well with the benchmark obtained from the HIES. Total household consumption and associated poverty rates tend not to differ statistically from the benchmark. When they do, they over-estimate consumption, but this may be because we have not fully accounted for all differences between the benchmark sample and the pilots. Also, the ranking of

⁴⁰ Such conditions do not hold: to provide an example, even the definition of household is defined differently across the Global MICS5 Pilot and the HIES 2010.

⁴¹ For instance, it is known that misreporting of consumption might vary by household characteristics, however it is difficult to disaggregate.

households based on total consumption is consistent with that obtained in HIES. However, individual components of consumption do not align well with the benchmark suggesting that such instruments are better at tracking overall consumption rather than its sub-components, which seems plausible given that these are shortened modules. It is thus feasible to include a shortened consumption module within the MICS. Pursuing such an activity opens up the possibility, first, of measuring absolute monetary welfare, and compiling poverty statistics and poverty profiles to improve understanding of monetary and non-monetary poverty; second, of improving the investigation of distributional issues over the wealth index; third, of comparing welfare across countries and over time; and finally, of increasing MICS usage for policy analysis and research purposes.

REFERENCES

- Bangladesh Bureau of Statistics (2011). Report of the Household Income and Expenditure Survey 2010, Bangladesh Bureau of Statistics, Dhaka.
- Beegle, K., J. De Weerd, J. Friedman, and J. Gibson (2010). "Methods of Household Consumption Measurement through Survey: Experimental Results from Tanzania", *World Bank Working Paper Series 5501*.
- Deaton, A. and M. Grosh (2000). 'Consumption', Chapter 5 in M. Grosh and P. Glewwe (eds.), *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*, Washington, DC: World Bank.
- Deaton, A. and S. Zaidi (2002). "Guidelines for Constructing Consumption Aggregates for Welfare Analysis", *World Bank LSMS Working Paper 135*.
- Elbers, C., J.O. Lanjouw and P. Lanjouw (2002). "Micro-Level Estimation of Welfare", *Policy Research Working Paper 2911*, World Bank: Washington D.C.
- Elbers, C., J.O. Lanjouw and P. Lanjouw (2003). "Micro-Level Estimation of Poverty and Inequality", *Econometrica*, 71-1: 355-364
- Freedman, D.A. (2008). "On Regression Adjustments to Experimental Data", *Advances in Applied Mathematics* 40(2): 180-193.
- Howe, L.D., J.R. Hargreaves, and S.R. Huttly (2008). "Issues in the Construction of Wealth Indices for the Measurement of Socio-economic Position in Low-income Countries", *Emerging Themes in Epidemiology* 2008 (5): 3.
- Kijima, Y. and P. Lanjouw (2003). "Poverty in India During the 1990s: A Regional Perspective", *Policy Research Working Paper No 3141*, The World Bank, Washington D.C.
- Moratti, M. and L. Natali (2012). "Measuring Household Welfare: Short versus long consumption modules," *Innocenti Working Paper 2012-04*, UNICEF Innocenti Research Centre.
- Pradhan, M. (2001). "Welfare Analysis with a Proxy Consumption Measure: Evidence from a repeated experiment in Indonesia", Economic and Social Institute, Free University, Amsterdam.
- Rutstein S.O. and K. Johnson (2004). "The DHS Wealth Index", *DHS Comparative Reports 6*, MEASURE DHS+ Calverton, MD.
- Simler, K. R., S. Mukherjee, G. Dava and G. Datt (2003). Rebuilding after War: Micro-level determinants of poverty reduction in Mozambique, International Food Policy Research Institute (IFPRI).
- World Bank (nd). Measuring Living Standards: Household Consumption and Wealth Indices. Quantitative Techniques for Health Equity Analysis. Technical Note #4. Available at: http://siteresources.worldbank.org/INTPAH/Resources/Publications/Quantitative-Techniques/health_eq_tn04.pdf