

Technical appendix

Strengthening accountability through media in Tanzania: final evaluation

April 2017

Research and Learning

Contents

Introduction	3
1. Survey sampling methodology.....	4
2. Regression analysis.....	5
Variables	5
Significance testing	5
Analysis	5
Full model results	7
Knowledge.....	7
Discussion.....	9
Participation	11

Introduction

This technical appendix is intended to be read alongside the associated BBC Media Action report, *Strengthening accountability through media in Tanzania: final evaluation*. This is available from: <http://dataportal.bbcmmediaaction.org/site/assets/uploads/2016/09/Tanzania-Country-Report.pdf>

I. Survey sampling methodology

As outlined in the table below, BBC Media Action carried out three national surveys.

Table I: Quantitative surveys conducted (2013-2016)

Study	Data collection	Sample size	Criteria
Midline survey	July 2013	n=4,043	Nationally representative, adults 15+
Tracker survey (omnibus)	February 2014	n=2,000	Nationally representative, adults 18+
Tracker survey (omnibus)	March 2016	n=2,000	Nationally representative, adults 15+

For each of these surveys, the sample was stratified across the major geographical divisions of the country (by province/region/state). Within these geographical divisions a probability proportional to size multistage cluster sample was employed. At all stages, the selection of clusters was random and self-weighting. The sampling frame for these surveys was constructed using the 2012 Tanzania Census. Within enumeration areas, predefined random starting points were used to begin household selection. Random walk was applied with a fixed household interval. Within households, a Kish grid was used to select respondents.

Data collection was carried out using face-to-face interviews and recorded using either paper and pen, or computer-assisted personal interviewing (CAPI).

After data collection, the samples were compared to the latest census data and, where necessary to correct for any imbalances in region/state, gender, age and location (urban vs rural), nested weights applied.

2. Regression analysis

This technical appendix summarises the results of analysis BBC Media Action carried out on the Tanzania midline dataset (2013), which was representative of Tanzania's national adult population (15+).

BBC Media Action carried out regression analysis to test the association between exposure to *Haba na Haba* and several governance outcomes (political knowledge, discussion and participation), while controlling for some of the other key factors which may influence these outcomes.

Variables

The independent variable for regression analysis was exposure to *Haba na Haba* (Little by Little), with two categories: regularly exposed to *Haba na Haba* (exposed to at least every other episode), and never exposed to *Haba na Haba*. Those who had been exposed, but not regularly, and those without access to media were set as missing.

The dependent variables were constructed as either categorical or continuous variables, dependent on the distribution of the outcome variables. Logistic regression was carried out for categorical dependent variables, and linear (OLS) regression was conducted for continuous dependent variables.

In addition to being based on past research and the specific country context, the confounders used in the analysis were chosen because they were hypothesised to be key factors in influencing the outcome variable. They therefore varied slightly across models.

Significance testing

Before carrying out regression analysis, BBC Media Action conducted statistical tests in order to measure the strength and the direction of bivariate relationships, as well as to test their significance. More precisely, BBC Media Action analysed:

- The relationship between the main independent variable (exposure) and the construct variables defined as outcomes (political knowledge, discussion and participation)
- The relationships among outcome variables
- The relationship between exposure and all the socio-demographic variables potentially associated with it (referred to as “confounders”)
- The relationship between the outcome variables and confounders

BBC Media Action conducted different types of significance tests according to the nature of the variables considered. T-tests and Mann-Whitney U-tests were used to compare the differences between means, Pearson's R and Spearman's Rho tests were used to ascertain correlation, and Chi-squared tests were conducted to measure associations. All significance tests were conducted with significance at the $p = 0.05$ level.

Analysis

As mentioned above, BBC Media Action carried out different types of regression analysis based on the dependent variable.

Logistic regression model: this allows researchers to work with categorical variables such as the binary variables where the distribution of the variables does not follow a normal and linear distribution that could have fitted better in another statistical model such as a linear regression. The logistic regression produces a probability value or odds ratio (OR) that indicates how much more likely it is that cases with specific attributes will fit into a model that explains the presence of certain outcomes. The regressions are calculated with a certain degree of confidence specified by the model. This confidence interval is used to

understand if the changes in one variable are associated with changes in the other as a result of a statistical relationship that can be explained by the model. Here, any value above 95% is considered as statistically significant.

The ordinary least squares (OLS) model: this allows researchers to work with a continuous dependent variable, derived through confirmatory factor analysis, and independent variables that have either continuous or categorical values. The regression coefficient for the independent variable provides key information indicating the estimated change in the dependent variable associated with a one unit increase in the independent variable. The model seeks to summarise this association by fitting a straight line to predict the value of the dependent variable based on the observed values of the independent variables.

BBC Media Action’s data satisfied the principle assumptions required for justifying the use of OLS: the relationships between the dependent and independent variables were linear and additive, and the error terms were normally distributed, constant, and were not correlated. With these assumptions met, a confidence interval for the regression line was calculated for each estimate and BBC Media Action was able to test whether the hypothesis of a zero slope – that is of no relationship between the two key variables of interest – existed in the true population.

Prior to analysis, BBC Media Action adopted the conventional standard of rejecting the null hypothesis at the 0.05 level. Given this, BBC Media Action expects that any estimated effects that are significantly associated with exposure to the programme of interest fall within the range reported in the confidence intervals 95% of the time.

Table 2: Overview of regression models

Model	Dataset	Model performance			Association with exposure		
		Sample size	R square	Significance (OLS only)	Association (OR/ coefficient)	95% confidence interval/ standard error (SE)	Significance
Regression 1: Knowledge OLS model	Midline	3,620	0.115	<0.001	0.347	0.090- 0.604	0.008
Regression 2: Discussion OLS model	Midline	3,627	0.102	<0.001	0.385	0.061- 0.709	0.020
Regression 3: Participated at least once versus never participated logistic model	Midline	3,800	0.183	-	1.966 (OR)	SE 0.137, Interval 1.380- 2.800	<0.001

Full model results

Knowledge

Table 3: Regression 1 – Knowledge OLS model

Dependent variable: average self-reported knowledge (0 to 10)

Predictor	Unstandardized coefficients		Standardized coefficient	Significance	95.0% confidence interval for B	
	Beta	Standard error	Beta		Lower bound	Upper bound
Not exposed to <i>Haba na Haba</i>	Ref	-	-	-	-	-
Regularly exposed to <i>Haba na Haba</i>	.347	.131	.042	.008	.090	.604
Group membership – not an active member	Ref	-	-	-	-	-
Active member	.195	.080	.040	.015	.037	.352
Interest in politics – not at all interested	Ref	-	-	-	-	-
Not very interested	.017	.154	.003	.912	-.285	.319
Somewhat interested	.590	.144	.107	.000	.308	.872
Very interested	.887	.136	.180	.000	.621	1.154
Region - Lake	Ref	-	-	-	-	-
Central	.508	.158	.056	.001	.198	.818
Eastern/Coastal	.076	.125	.012	.545	-.170	.321
Western	.161	.136	.021	.238	-.106	.429
Northern	.255	.124	.037	.039	.012	.498
Southern	.137	.192	.012	.476	-.240	.514
Southern Highlands	.582	.127	.083	.000	.332	.832
Zanzibar	-.438	.252	-.029	.083	-.933	.057
Male	Ref	-	-	-	-	-
Female	-.110	.079	-.022	.166	-.265	.045
Rural	Ref	-	-	-	-	-
Urban	.319	.094	.058	.001	.135	.504
Age 15-24	Ref	-	-	-	-	-
Age 25-34	.489	.111	.090	.000	.272	.707
Age 35-44	.593	.134	.092	.000	.330	.855
Age 45-54	.513	.156	.065	.001	.207	.819
Age 55-64	.686	.190	.066	.000	.313	1.058
Age 65+	.321	.225	.026	.154	-.121	.763
Education - illiterate	Ref	-	-	-	-	-
Literate	.922	.188	.119	.000	.553	1.291
Completed primary	1.197	.162	.243	.000	.880	1.514
Completed secondary	1.599	.175	.286	.000	1.256	1.942
Completed college/university	2.462	.220	.253	.000	2.032	2.893
Income - medium	Ref	-	-	-	-	-
Low	-.300	.101	-.049	.003	-.497	-.103

High	.127	.137	.015	.356	-.143	.397
Marital status – married, living with spouse	Ref	-	-	-	-	-
Single	-.241	.108	-.047	.026	-.453	-.029
Married, not living with spouse	.208	.219	.015	.341	-.221	.638
Divorced/separated	-.168	.174	-.015	.333	-.508	.172
Widowed	.038	.177	.004	.831	-.309	.385
In a marriage where the husband has more than one wife	-.041	.488	-.001	.933	-.997	.916
Constant	2.123	.234	-	.000	1.664	2.582

The model had an adjusted R square of 0.115. The Durbin-Watson value was 1.667. The F statistic was 16.645 (significance < 0.001).

Discussion

Table 4: Regression 2 – Discussion OLS model

Dependent variable: average discussion (0 to 10)

Predictor	Unstandardized coefficients		Standardized coefficient	Significance	95.0% confidence interval for B	
	Beta	Standard error	Beta		Lower bound	Upper bound
Not exposed to <i>Haba na Haba</i>	Ref	-	-	-	-	-
Regularly exposed to <i>Haba na Haba</i>	.385	.165	.037	.020	.061	.709
Group membership – not an active member	Ref	-	-	-	-	-
Active member	.729	.101	.118	.000	.530	.927
Interest in politics – not at all interested	Ref	-	-	-	-	-
Not very interested	.493	.194	.059	.011	.112	.875
Somewhat interested	1.241	.181	.179	.000	.886	1.597
Very interested	1.357	.171	.220	.000	1.021	1.693
Region - Lake	Ref	-	-	-	-	-
Central	-.692	.200	-.061	.001	-1.084	-.301
Eastern/Coastal	-.443	.158	-.054	.005	-.753	-.134
Western	.013	.172	.001	.941	-.325	.350
Northern	-.017	.156	-.002	.914	-.323	.289
Southern	.074	.243	.005	.760	-.401	.550
Southern Highlands	-.591	.161	-.067	.000	-.906	-.276
Zanzibar	-.994	.318	-.052	.002	-1.618	-.369
Male	Ref	-	-	-	-	-
Female	-.565	.100	-.091	.000	-.760	-.369
Rural	Ref	-	-	-	-	-
Urban	-.112	.119	-.016	.347	-.344	.121
Age 15-24	Ref	-	-	-	-	-
Age 25-34	.270	.140	.040	.054	-.004	.544
Age 35-44	.685	.169	.085	.000	.354	1.016
Age 45-54	.446	.197	.045	.023	.060	.833
Age 55-64	.652	.239	.050	.007	.182	1.121
Age 65+	.447	.284	.029	.116	-.110	1.004
Education - illiterate	Ref	-	-	-	-	-
Literate	.315	.237	.032	.184	-.150	.781
Completed primary	.517	.204	.084	.011	.117	.916
Completed secondary	1.038	.221	.149	.000	.606	1.470
Completed college/university	1.510	.277	.124	.000	.967	2.052
Income - medium	Ref	-	-	-	-	-
Low	-.615	.127	-.080	.000	-.864	-.366
High	-.107	.173	-.010	.537	-.447	.233

Marital status – married, living with spouse	Ref	-	-	-	-	-
Single	-.364	.137	-.057	.008	-.632	-.096
Married, not living with spouse	.052	.276	.003	.851	-.490	.594
Divorced/separated	-.453	.219	-.033	.039	-.882	-.023
Widowed	-.649	.223	-.050	.004	-1.086	-.212
In a marriage where the husband has more than one wife	.845	.616	.022	.170	-.362	2.052
Constant	3.248	.295	-	.000	2.669	3.828

The model had an adjusted R square of 0.102. The Durbin-Watson value was 1.663. The F statistic was 14.549 (significance < 0.001).

Participation

Table 5: Regression 3 – Participated at least once versus never participated logistic model

Variable	Standard error	Significance level	Odds ratio
Not exposed to <i>Haba na Haba</i>	Ref	-	-
Regularly exposed to <i>Haba na Haba</i>	.137	.000	1.966
Group membership – not an active member	Ref	-	-
Active member	.073	.000	2.163
Interest in politics – not at all interested	Ref	.014	-
Not very interested	.145	.710	1.055
Somewhat interested	.134	.481	1.099
Very interested	.127	.020	1.346
Male	Ref	-	-
Female	.075	.002	.795
Rural	Ref	-	-
Urban	.084	.203	.899
Age 15-24	Ref	.000	-
Age 25-34	.101	.000	1.826
Age 35-44	.125	.000	2.617
Age 45-54	.151	.000	3.104
Age 55-64	.209	.000	5.362
Age 65+	.253	.000	5.466
Education – no schooling	Ref	.000	-
Some primary	.190	.052	1.446
Completed primary	.166	.000	2.008
Completed secondary	.177	.000	2.058
Completed college/university	.222	.000	3.332
Income – low	Ref	.003	-
Medium	.098	.005	1.322
High	.126	.130	.827
Marital status – single	Ref	.026	-
Married, living with spouse	.099	.122	.857
Married, not living with spouse	.226	.176	1.358
Divorced/separated	.165	.003	.610
Widowed	.183	.914	1.020
Living with partner	.467	.559	.761
In a marriage where the husband has more than one wife	.358	.327	.704
Constant	.225	.000	.251

The Nagelkerke R statistic for this model was .183. The Hosmer and Lemeshow statistic had a chi-square of 4.718 and a significance level of .787.