Supporting Information

Predictors of drinking water boiling and bottled water consumption in rural China: A hierarchical modeling approach

Alasdair Cohen[1,2]*, Qi Zhang[3], Qing Luo[3], Yong Tao[3], John M. Colford, Jr.[2], Isha Ray[4]

1. Department of Environmental Science, Policy and Management, University of California at Berkeley, Berkeley, California, United States of America

2. School of Public Health, University of California at Berkeley, Berkeley, California, United States of America

3. National Center for Rural Water Supply Technical Guidance, Chinese Center for Disease Control and Prevention, Beijing, China

4. Energy and Resources Group, University of California at Berkeley, Berkeley, California, United States of America

*Corresponding author: Alasdair Cohen; Department of Environmental Science, Policy, and Management; University of California, Berkeley; 130 Mulford Hall #3114; Berkeley, CA 94720-3114; Telephone: 510 643 4554; Fax: 510 643 5438; E-mail: alasdair.cohen@linacre.oxon.org

Supporting Information:

Pages: 25
Figures: 11
Tables: 18

1. Completed STROBE Checklist 2

2. Methods: MPAT subcomponent and component names & organization 4

3. Results: Process used to identify MPAT survey items & covariates 6

4. Results: MPAT histograms, correlation matrix, and values..... 10

5. Results: Covariates blocks and adjustments for each model 12

6. Results: Sensitivity analyses for all models 15

7. Results: Demographic, socioeconomic, and water-related data 18

8. Discussion: Key factors associated with HWT and bottled water use 23

1. Completed STROBE Checklist

Strengthening the reporting of observational studies in epidemiology (STROBE)

Checklist for cross-sectional studies (v4)

Source: http://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_case-control.doc

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8 & SI
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8 & SI
Bias	9	Describe any efforts to address potential sources of bias	6-8 & SI
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-8 & SI
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9 & SI
		(*b*) Describe any methods used to examine subgroups and interactions	8-9 & SI
		(*c*) Explain how missing data were addressed	9
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	8
		(*e*) Describe any sensitivity analyses	9 & SI
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6 & 10
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic,	10

		clinical, social) and information on exposures and potential confounders	& SI
		(b) Indicate number of participants with missing data for each variable of interest	10
Outcome data	15*	Report numbers of outcome events or summary measures	10 & SI
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11-15 & SI
		(*b*) Report category boundaries when continuous variables were categorized	15 & SI
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	SI
Discussion			
Key results	18	Summarise key results with reference to study objectives	16-19
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	17-18
Generalisability	21	Discuss the generalisability (external validity) of the study results	17-18
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1

Note: Page numbers reference the submitted manuscript and may not align with the published version.

2. Methods: MPAT subcomponent and component names & organization

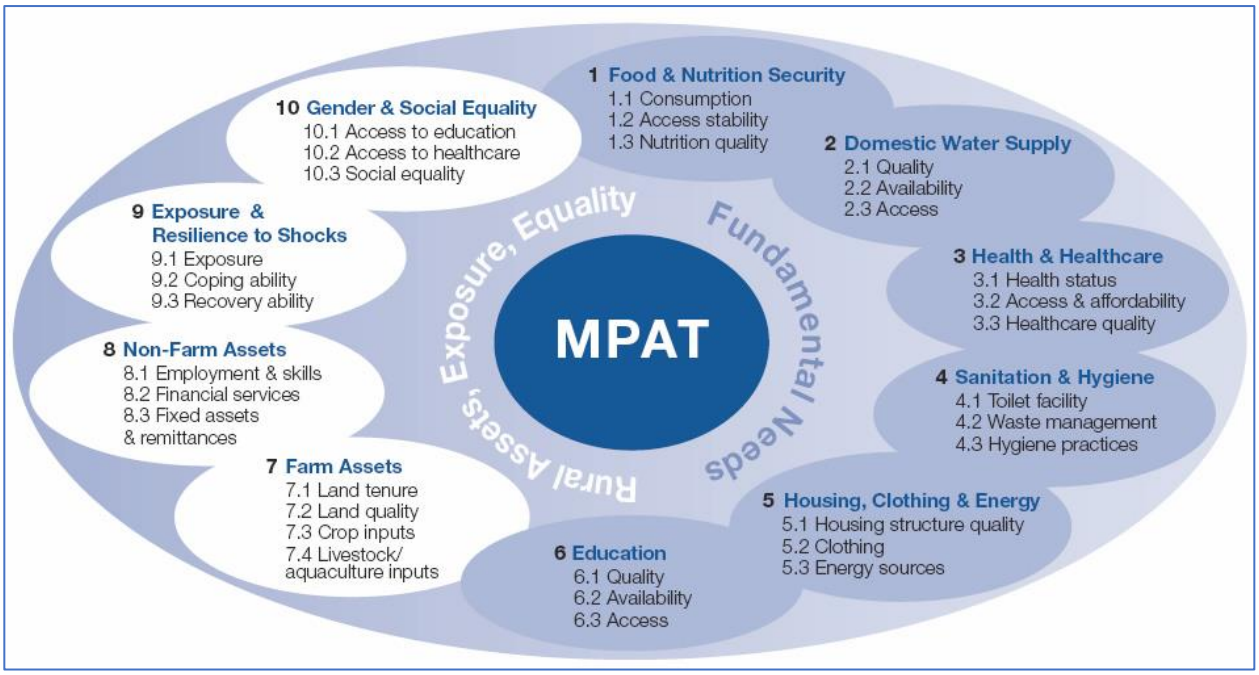

Figure S1: MPAT Components & Subcomponents (source: https://www.ifad.org/topic/mpat/faq/tags/mpat)

A more detailed figure, with the MPAT survey questions associated with each subcomponent, is available online:
https://www.ifad.org/documents/10180/15a82afc-c024-4354-ac8a-0dbc555f4f44

The MPAT User's Guide (2014) which provides details with regard to the weightings and cardinal values used to aggregate the MPAT indicators is available online:
https://www.ifad.org/documents/10180/8f51cc65-88f8-49cb-926b-236aecc32734

The MPAT Excel Spreadsheet is available online:
https://www.ifad.org/documents/10180/45ba3297-4784-402c-8251-a2be6902aee1

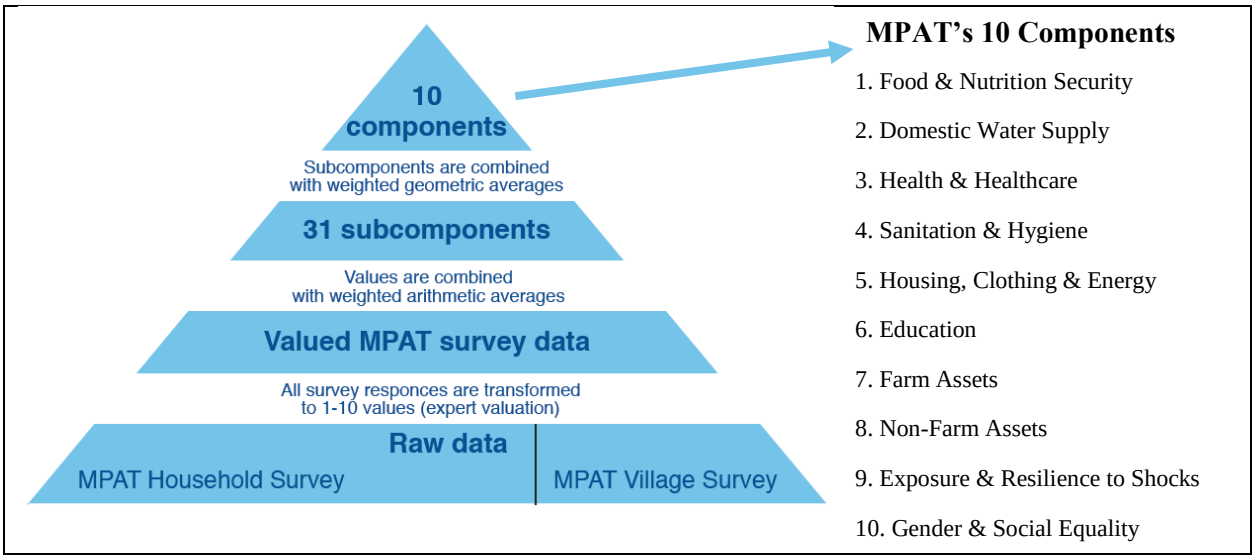

Figure S2: MPAT indicator aggregation scheme and main components

Figure reproduced from: IFAD, *The Multidimensional Poverty Assessment Tool: User's Guide*. The International Fund for Agricultural Development: Rome, 2014.

3. Results: Process used to identify MPAT survey items & covariates

Table S1: MPAT components, subcomponents, & survey items identified by stepwise regression

Step 1		**Step 2**		**Step 3**	
MPAT Components Identified (at p<0.2)		MPAT Subcomponents Identified (at p<0.15)		MPAT Survey Items Identified (at p<0.15)	
Boil/Un.	Bottled/Un.	Boil/Un.	Bottled/Un.	Boil/Un.	Bottled/Un.
2	2	2.1	2.1	34	32 & 34
3		3.2		13 & 14	11 & 14
5	5	5.1	5.1 & 5.3	21	17* & 21
6.3*		---	---	---	---
8		---	---	---	---
10		10.2		---	---

Note: Two models were used for each step to identify variables potentially associated with boiling and bottled water (in part because bottled water is not usually considered a form of HWT) using binary dependent variables: Boil-vs-Untreated and Bottled-vs-Untreated.

*Subcomponent 6.3 was used in place of component six because there was no available data for subcomponents 6.1 and 6.2 (due to survey item censorship).

In the first stage of the stepwise regression process five components were significantly associated with boiling: component numbers 2, 3, 5, 6.3 (used in place of 6 due to censored questions from subcomponents 6.1 and 6.2), 8, and 10. Stage two identified five subcomponents and among these seven survey items were identified in stage three (see Table S2). For Q17, 83% of HHs in County A have walls made of cement blocks which is assigned a value of 10 (the best value) under MPAT and 65% of HHs in County B have brick walls which is assigned a value of 8 under MPAT. Thus, when Q17 is converted into a binary, 98% of all households have structurally sound walls (i.e., the distinction between a score of 8 and 10 is lost since both fall under the value of one for the new dummy variable). As such, the other survey item from subcomponent 5.1 – the home's ability to withstand extreme weather (Q19) - was used to represent this construct. Q34 (whether the household treats their water or not) was not used for the modeling due to obvious collinearity issues with the outcomes.

Table S2: MPAT survey questions identified via stepwise logistic regression

MPAT Subcomponent Name (MPAT Component Name)	**Associated survey questions [and MPAT aggregation weights]**
2.1 Quality (Domestic water supply)	32) What is the **primary source** (meaning the source that water comes from immediately before being used) of the water your household uses for drinking and cooking inside the home? [45%] *34) Does your household treat water before drinking it (any treatment method: boiling, allowing to settle, filter, chemical treatment, etc.)? [35%]
3.2 Access & affordability (Health and health care)	11) How much time does it take for members of your household to reach the nearest health centre that can diagnose simple illness, or treat simple injuries and prescribe basic medicines? [25% or 38.5%] 13) How much time does it take for members of your household to reach the nearest health centre that can diagnose and treat complicated or serious illnesses or injuries (can perform surgery)? [35% or 0%] 14) Can your household afford professional treatment for serious illness or injury? [40% or 61.5%]
5.1 Housing structure quality (Housing, clothing & energy)	**17) What is the **primary** construction material of the housing unit's exterior walls? [70%] **19) Can your home withstand strong winds, severe rain, snow or hail without significant damage [30%]
5.3 Energy sources (Housing, clothing & energy)	21) What is the **primary** fuel source your household uses for cooking? [40% or 57%]

Notes:

* Q34 (whether the household treats their water or not) was not used for the modeling due to obvious collinearity issues with the outcomes.
**For Q17, 83% of HHs in County A had walls made of cement blocks which is assigned a value of 10 (the best value) under MPAT and 65% of HHs in County B had brick walls which is assigned a value of 8 under MPAT. Thus, when Q17 was converted into a binary, 98% of all households had structurally sound walls (i.e., the distinction between a score of 8 and 10 was lost since both fell under the value of one for the new dummy variable). Therefore, the other survey item from subcomponent 5.1 – the home's ability to withstand extreme weather (Q19) - was used to represent this construct instead.

Table S3: Covariates assessed for associations with HWT methods

Variable	Type	Definition
Water & behavior-related		
HH believes DWQ is good/very good	D	1= HH perceives their drinking water quality to be good or very good, 0= other
HH has improved drinking water source	D	1= "Improved" water source, 0= Unimproved
HH believes most/all nearby relatives boil	D	1= "most" or "all" of relatives boil, 0=other
HH believes most/all neighbors boil	D	1= "most" or "all" of relatives boil, 0=other
Access to health services		
Minutes to clinic for basic care	C	Minutes to reach nearest health clinic
Minutes to clinic for advanced care	C	Minutes to reach nearest health center that can address serious illness or injury
HH can afford professional care	D	1= HH can afford professional medical care, 0= other
Economic indicators		
Number of TVs in HH / HH population	C	Number of TVs in HH/HH population
Village-average price for 19L W bottle	~C	Average cost of a 19L bottle of water per village
Home can withstand severe weather	D	1= Home can withstand severe weather, 0= Home cannot withstand severe weather
HH uses safe fuel for cooking & heating	D	1= HH uses safe fuel (no/low HAP potential), 0= other
Demographic & gender-related		
Head of the HH's age	C	Head of the HH's age
Male-headed HHs*	D	1=Male, 0=Female or Joint (Female & Male)
Married head of HH*	D	1= head of HH is married, 0=other
Head of the HH is a single female** Head of the HH is a single male** Head of the HH is a married female or F&M** (Reference) Head of the HH is a married male**	I	1= Single F head of HH, 0= other 1= Single M head of HH, 0= other 1= Married F or F&M head of HH, 0= other Married M head of HH (all three variables = 0)
Head of the HH is literate	D	1= head of the HH is literate, 0= illiterate
HH population (live in HH >9 months)	C	Adults and children living in the HH

HH = Household | D = Dichotomous | C = Continuous | I = Interaction terms
Variables marked with "**" are interaction terms that are used for some models instead of variables marked "*"
MPAT-derived variables in *italics*

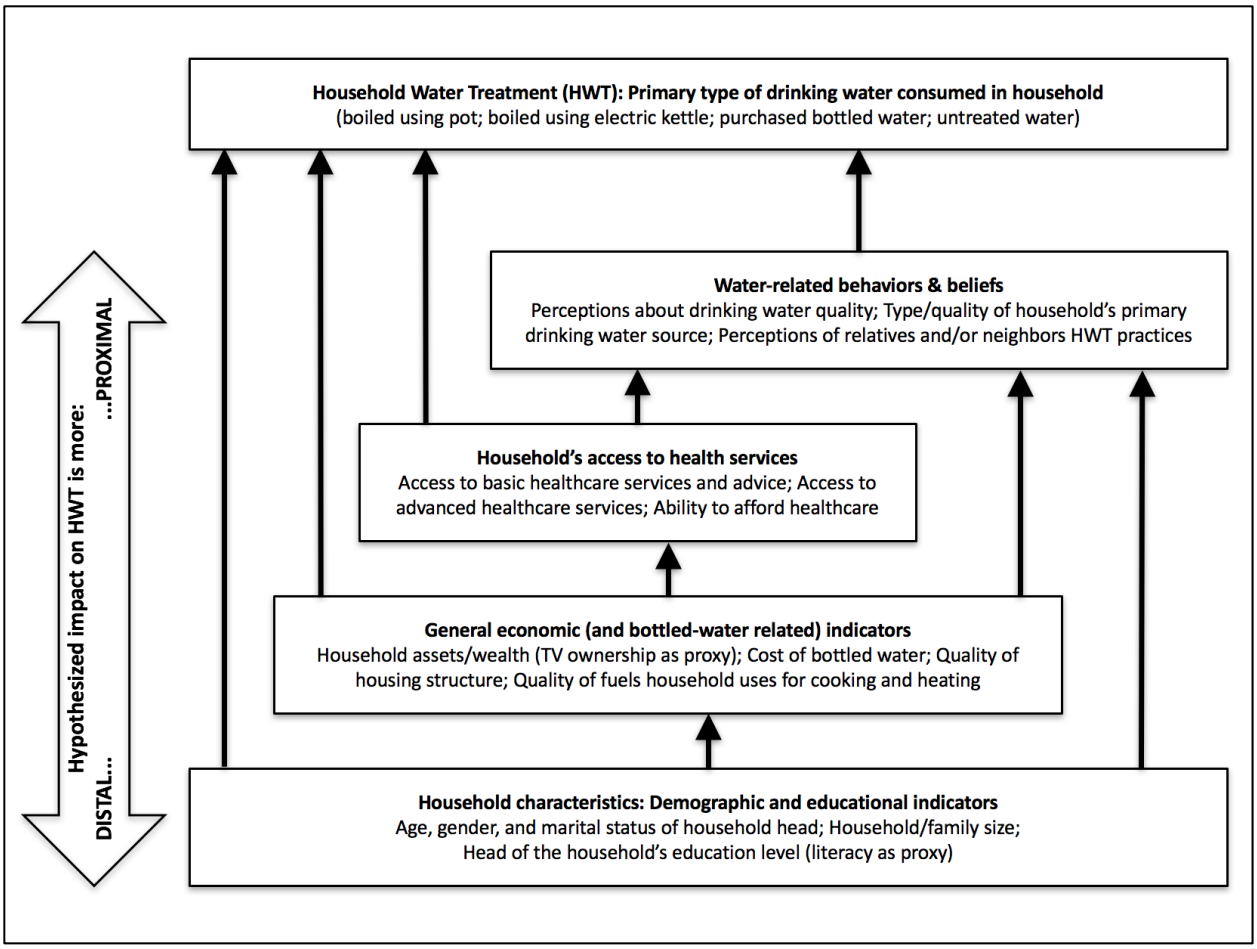

Figure S3: Simplified hierarchical conceptual framework of factors hypothesized to impact HWT use (grouped by the thematic blocks used for model construction)

4. Results: MPAT histograms, correlation matrix, and values

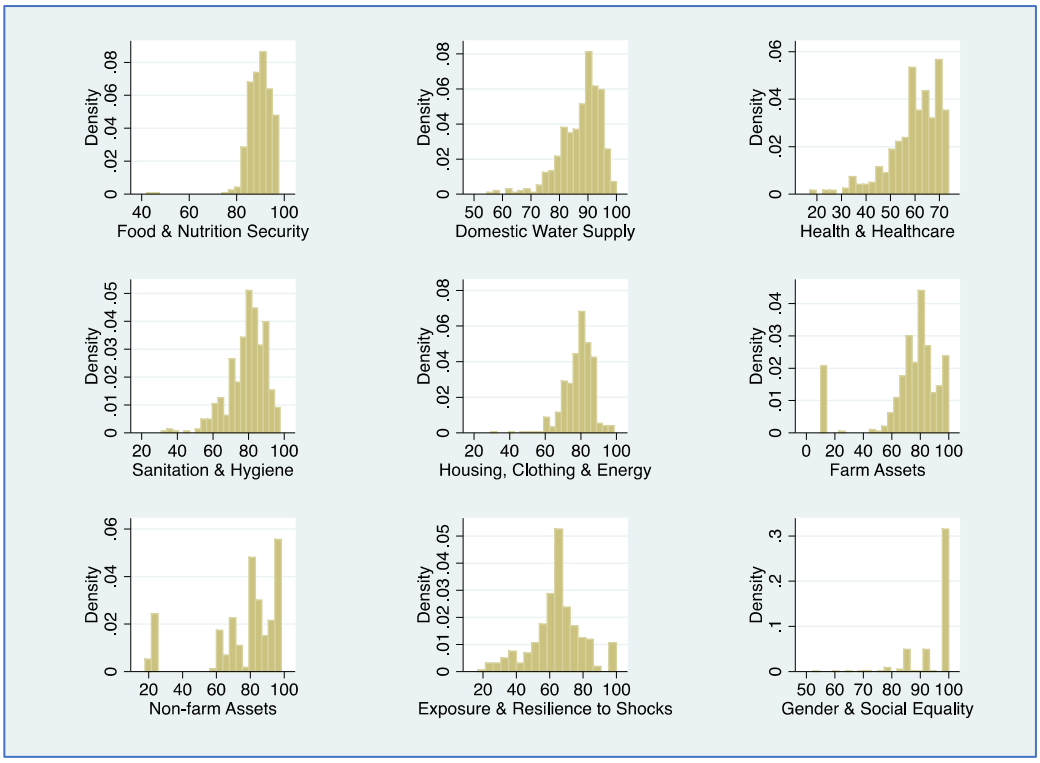

Figure S4: Matrix of MPAT component histograms

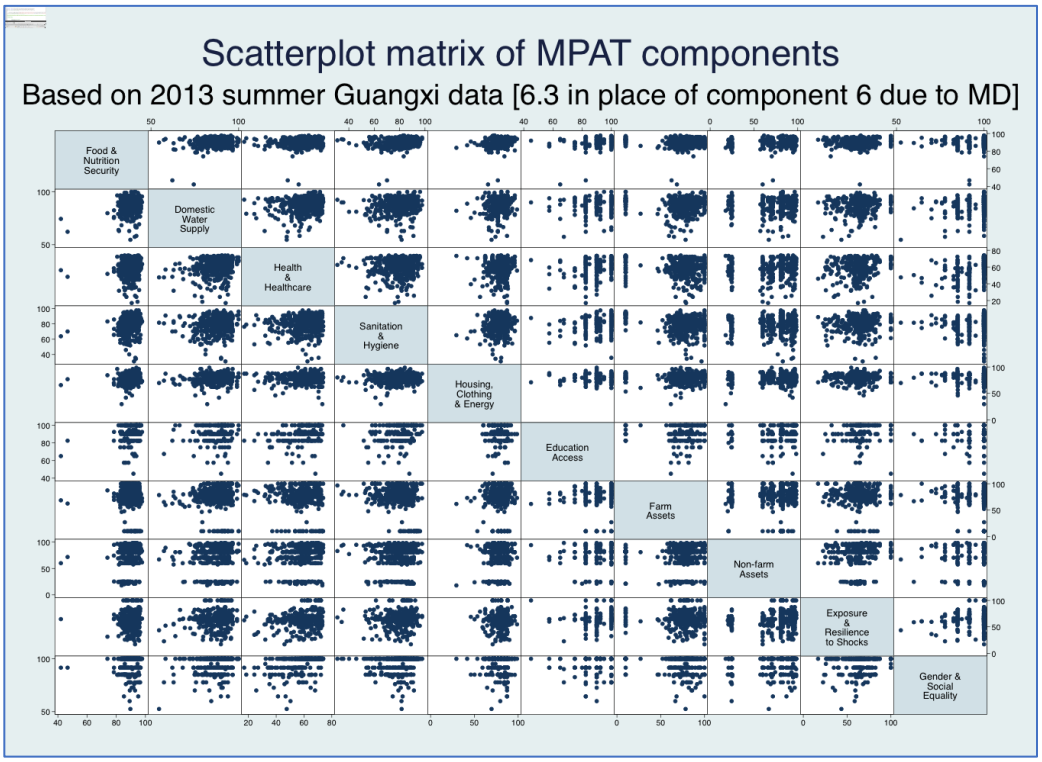

Figure S5: Scatterplot matrix of MPAT component values

Note for Figure S5: One method for evaluating the overall reliability of the MPAT indicators is to examine the correlations between components. With regard to indicator construction generally, each component should represent and measure a different construct (and no data should be used for multiple components – i.e., no double-counting). Thus, no pair of MPAT components should be highly correlated since the data upon which each component is based, while related to data in other components, measures a specific construct.

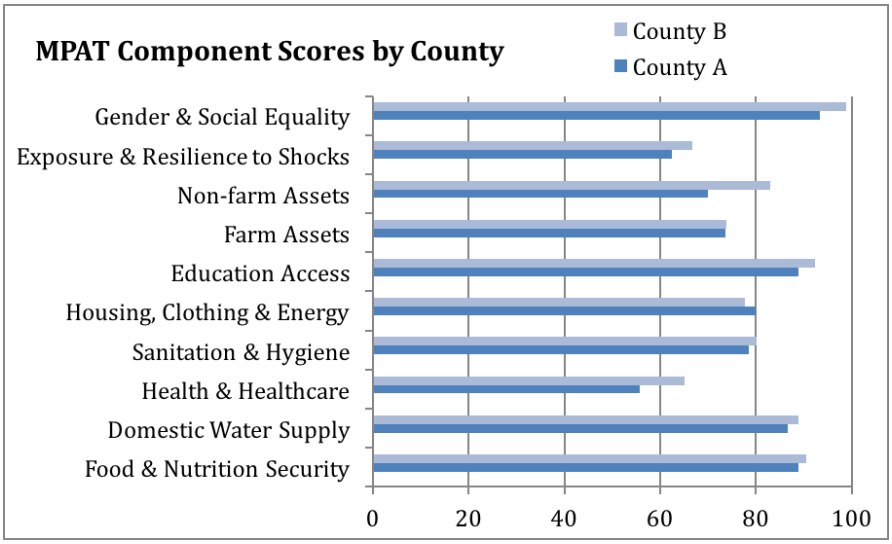

Figure S6: MPAT Components values by county

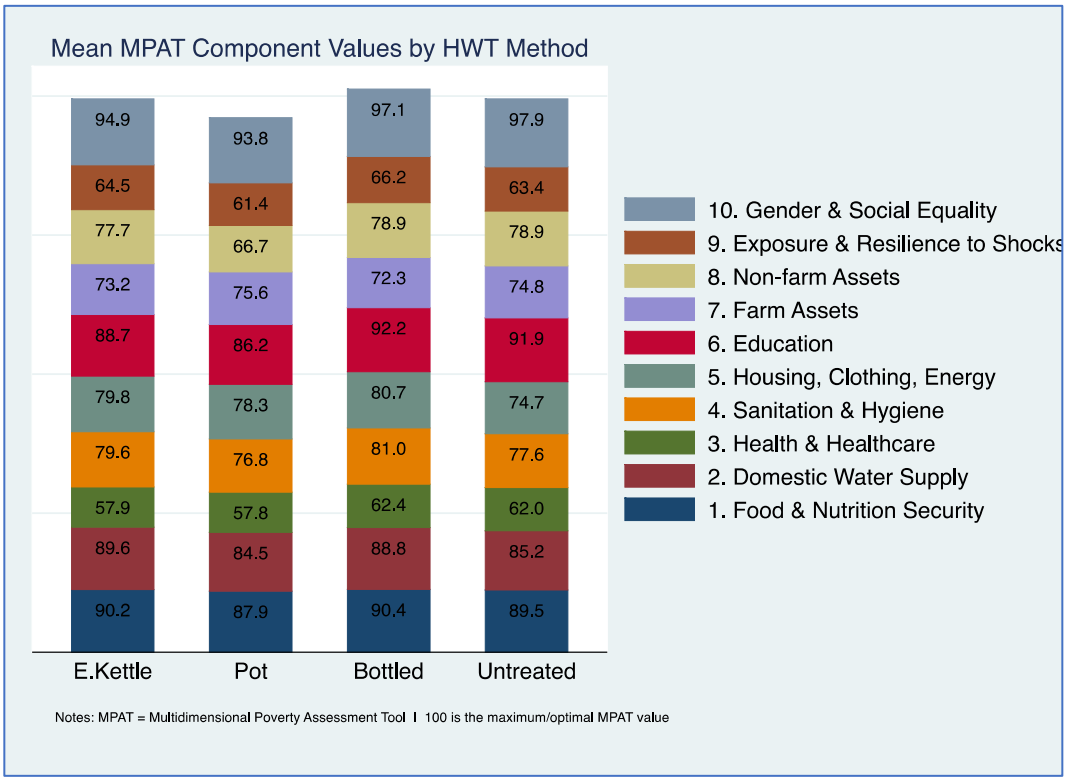

Figure S7: MPAT Components values by HWT Method

5. Results: Covariates blocks and adjustments for each model

Table S4: Boil-vs-Untreated: Model results by covariate block & final model

	Risk Ratio (95% CI)			
				Final Model
Water & behavior-related				
Believe DWQ is good/very good	0.75** (0.61-0.91)	0.78* (0.64-0.95)	0.79* (0.65-0.97)	0.78** (0.64-0.94)
Improved drinking water source	1.11 (0.91-1.34)	1.09 (0.92-1.29)	1.09 (0.93-1.27)	1.10 (0.93-1.29)
Access to health services				
Minutes to clinic for basic care		1.005** (1.002-1.008)	1.005** (1.002-1.009)	1.005** (1.002-1.008)
HH can afford professional care		0.90[a] (0.79-1.01)	0.91 (0.79-1.05)	0.91 (0.79-1.04)
Economic indicators				
TVs in HH / HH population			0.86 (0.70-1.05)	0.83[b] (0.68-1.02)
Home can withstand severe weather			1.12 (0.80-1.54)	1.12 (0.80-1.56)
Demographic & gender-related				
Head of the HH's age				1.003 (0.997-1.010)
HH head is married F or F&M				1.08 (0.87-1.35)
HH head is single male				0.95 (0.72-1.26)
HH head is single female				1.36** (1.12-1.66)
HH head is literate				1.05 (0.88-1.26)
HH population (live in HH >9 months)				0.997 (0.967-1.028)
Model indicators				
Log pseudo-likelihood	-259.2	-254.14	-237.51	-230.82
n	273	269	252	246

HH=household | * p<0.05; ** p<0.01; *** p<0.001 | [a] p=0.071 | [b] p=.076

Table S5: Kettle-vs-Pot: Model results by covariate block & final model

	Risk Ratio (95% CI)				
				Full Model	**Final Model**
Water & behavior-related					
Believe DWQ is good/very good	1.23 (.81-1.88)	1.20 (.86-1.68)	1.22 (.87-1.70)	1.13 (.82-1.57)	1.12 (0.81-1.56)
Improved drinking water source	0.98 (0.58-1.65)	0.93 (0.58-1.50)	0.92 (0.56-1.50)	0.98 (0.63-1.52)	0.99 (0.64-1.54)
Access to health services					
Minutes to clinic for basic care		0.992 (0.967-1.017)	0.992 (0.964-1.020)	0.996 (0.972-1.021)	0.997 (0.974-1.021)
HH can afford professional care		0.96 (0.71-1.30)	0.91 (0.69-1.21)	0.94 (0.68-1.30)	---
Economic indicators					
TVs in HH / HH population			1.16 (0.90-1.50)	1.43** (1.20-1.75)	1.42** (1.16-1.74)
Home can withstand severe weather			1.25 (0.71-2.23)	1.07 (0.63-1.83)	1.04 (0.57-1.89)
Demographic & gender-related					
Head of the HH's age				0.990* (0.990-0.999)	0.989* (0.980-0.998)
HH head is male				1.14 (0.74-1.76)	1.14 (0.74-1.75)
HH head is married				1.42 (0.72-2.80)	1.43 (0.73-2.80)
HH head is literate				1.18 (0.82-1.69)	1.16 (0.83-1.63)
HH population (live in HH >9 months)				1.13*** (1.08-1.19)	1.13*** (1.07-1.19)
Model indicators					
Log pseudo-likelihood	-176.89	-172.28	-160.55	-150.59	-150.63
n	199	195	182	177	177
HH=household | * p<0.05; ** p<0.01; *** p<0.001					

Table S6: Bottled-vs-Boil: Model results by covariate block & final model

	Risk Ratio (95% CI)				
				Full Model	**Final Model**
Water & behavior-related					
Believe DWQ is good/very good	1.19 (0.81-1.73)	1.16 (0.84-1.59)	1.03 (0.75-1.41)	0.99 (0.71-1.37)	---
Access to health services					
Minutes to clinic for advanced care		0.98** (0.97-0.99)	0.98*** (0.97-0.99)	0.98** (0.97-0.99)	0.98** (0.97-0.99)
HH can afford professional care		1.13 (0.83-1.53)	1.57** (1.15-2.14)	1.61** (1.19-2.17)	1.61** (1.18-2.19)
Economic indicators					
TVs in HH / HH population			0.88 (0.62-1.24)	0.97 (0.68-1.39)	---
Village-average price for 19L W bottle			1.06 (0.92-1.23)	1.06 (0.91-1.25)	1.06 (0.91-1.25)
Home can withstand severe weather			0.46** (0.29-0.73)	0.44*** (0.30-0.66)	0.49*** (0.35-0.67)
Demographic & gender-related					
Head of the HH's age				0.988** (0.98-0.995)	0.985*** (0.978-0.993)
HH head is male				1.02 (0.66-1.57)	0.995 (0.65-1.53)
HH head is married				0.97 (0.65-1.45)	0.93 (0.63-1.39)
HH head is literate				1.21 (0.93-1.57)	1.26* (1.002-1.58)
HH population (live in HH >9 months)				1.06 (0.99-1.13)	1.05 (0.99-1.12)
Model indicators					
Log pseudo-likelihood	-276.34	-268.46	-246.09	-228.50	-246.04
n	351	347	323	307	331

Notes: HH=household | * p<0.05; ** p<0.01; *** p<0.001

6. Results: Sensitivity analyses for all models

Table S7: Sensitivity analysis results for Boil/Untreated model

	Risk Ratio (95% CI)		Odds Ratio (95% CI)	
	Bootstrap (1,000x)	**Robust SE[a]**	**MLM Logit[b]**	**Logit**
Water & behavior-related				
Believe DWQ is good/very good	.78* (.63-.96)	.78* (.64-.94)	.40* (.19-.87)	.38** (.19-.77)
Improved drinking water source	1.10 (.90-1.33)	1.10 (.93-1.30)	1.88 (.76-4.67)	1.24 (.64-2.39)
Access to health services				
Minutes to clinic for basic care	1.00 (.99-1.01)	1.00* (1.00-1.01)	1.06 (.99-1.12)	1.04 (.99-1.08)
HH can afford professional care	.91 (.79-1.05)	.91 (.77-1.07)	.76 (.34-1.72)	.66 (.32-1.35)
Economic indicators				
TVs in HH / HH population	.83 (.66-1.04)	.83 (.66-1.04)	.54 (.27-1.06)	.52* (.27-.99)
Home can withstand severe weather	1.12 (.77-1.62)	1.12 (.82-1.52)	2.21 (.64-7.64)	1.79 (.60-5.26)
Demographic & gender-related				
Head of the HH's age	1.003 (.997-1.009)	1.003 (.997-1.010)	1.01 (.983-1.040)	1.02 (.989-1.042)
HH head is married F or F&M	1.08 (.84-1.4)	1.08 (.91-1.29)	2.03 (.45-9.12)	1.48 (.43-5.060)
HH head is single male	.95 (.24-3.8)	.95 (.64-1.4)	.58 (.11-3.10)	.71 (.16-3.24)
HH head is single female	1.36** (1.1-1.7)	1.36** (1.1-1.7)	8.89 (.83-95.54)	5.73 (.64-51.00)
HH head is literate	1.05 (.87-1.3)	1.05 (.90-1.23)	1.43 (.60-3.38)	1.30 (.60-2.83)
HH population	.997 (.963-1.03)	.997 (.956-1.039)	.975 (.818-1.162)	.989 (.839-1.165)
Model indicators				
Log pseudo-likelihood	-230.82	-230.82	-	-
Log likelihood	-	-	-124.37	-128.54
n	246	246	246	246

Notes: HH=household | | * p<0.05; ** p<0.01; *** p<0.001
[a] Not cluster-adjusted SE | [b] Mixed effects multi-level logit

Table S8: Sensitivity analysis results for Kettle/Pot model

	Risk Ratio (95% CI)		Odds Ratio (95% CI)	
	Bootstrap (1,000x)	**Robust SE[a]**	**MLM Logit[b]**	**Logit**
Water & behavior-related				
Believe DWQ is good/very good	1.12 (.77-1.65)	1.12 (.85-1.48)	.83 (.31-2.19)	1.31 (.59-2.91)
Improved drinking water source	.99 (.59-1.68)	.99 (.76-1.28)	.69 (.25-1.90)	.92 (.44-1.91)
Access to health services				
Minutes to clinic for basic care	.99 (.96-1.03)	.99 (.99-1.01)	1.01 (.97-1.04)	.99 (.967-1.02)
Economic indicators				
TVs in HH / HH population	1.42** (1.11-1.83)	1.42** (1.11-1.82)	3.67* (1.12-11.96)	3.16* (1.11-8.99)
Home can withstand severe weather	1.04 (.52-2.06)	1.04 (.61-1.75)	.83 (.20-3.50)	1.07 (.29-4.03)
Demographic & gender-related				
Head of the HH's age	.989* (.980-.999)	.989* (.979-.999)	.982 (.953-1.012)	.975 (.949-1.00)
HH head is male	1.14 (.68-1.9)	1.14 (.72-1.81)	1.22 (.45-3.29)	1.28 (.50-3.23)
HH head is married	1.43 (.06-35.5)	1.43 (.72-2.82)	2.18 (.57-8.28)	1.89 (.58-6.21)
HH head is literate	1.16 (.80-1.69)	1.16 (.83-1.62)	1.48 (.64-3.42)	1.49 (.70-3.20)
HH population	1.13*** (1.068-1.195)	1.13*** (1.058-1.207)	1.43** (1.147-1.793)	1.47*** (1.19-1.8)
Model indicators				
Log pseudo-likelihood	-150.63	-150.63	-	-
Log likelihood	-	-	-99.85	-103.48
n	177	177	177	177

Notes: HH=household | | * p<0.05; ** p<0.01; *** p<0.001
[a] Not cluster-adjusted SE | [b] Mixed effects multi-level logit

Table S9: Sensitivity analysis results for Bottled/Boil model

	Risk Ratio (95% CI)		Odds Ratio (95% CI)	
	Bootstrap (1,000x)	**Robust SE[a]**	**MLM Logit[b]**	**Logit**
Access to health services				
Minutes to clinic for advanced care	.98** (.97-.995)	.98*** (.97-.993)	.98*[c] (.95-.999)	.97** (.95-.988)
HH can afford professional care	1.61** (1.14-2.28)	1.61** (1.13-2.29)	2.55** (1.27-5.10)	2.30** (1.25-4.22)
Economic indicators				
Village-average price for 19L W bottle	1.07 (.87-1.30)	1.07 (.98-1.16)	1.21 (.78-1.89)	1.13 (.97-1.32)
Home can withstand severe weather	.49*** (.33-.71)	.49*** (.33-.71)	.16*** (.06-.39)	.22*** (.10-.50)
Demographic & gender-related				
Head of the HH's age	.985*** (.978-.993)	.985** (.975-.996)	.964** (.942-.986)	.971** (.952-.991)
HH head is male	.99 (.63-1.58)	.99 (.69-1.44)	.88 (.38-2.01)	1.01 (.48-2.12)
HH head is married	.93 (.61-1.43)	.93 (.61-1.42)	.96 (.35-2.62)	.89 (.38-2.11)
HH head is literate	1.26 (.97-1.13)	1.26 (.91-1.75)	2.03* (1.06-3.91)	1.53 (.86-2.74)
HH population	1.05 (.98-1.13)	1.05 (.99-2.52)	1.13 (.98-1.30)	1.10 (.98-1.24)
Model indicators				
Log pseudo-likelihood	-246.04	-246.04	-	-
Log likelihood	-	-	-181.33	-200.51
n	331	331	331	331

Notes: HH=household | | * p<0.05; ** p<0.01; *** p<0.001
[a] Not cluster-adjusted SE | [b] Mixed effects multi-level logit | [c] p=0.050

7. Results: Demographic, socioeconomic, and water-related data

Table S10: Summary statistics for key covariates by head of household gender

HoHH Gender		HoHH Age	HoHH Marital	HoHH Literacy	Min. to Basic Health Service	Min to Adv. Health Service	Able to Afford Pro. Health Care	TVs by HH pop.
Male	**Mean**	52.10	0.96	0.75	10.47	25.12	0.60	0.59
	SD	12.12	0.21	0.43	10.20	13.76	0.49	0.43
	n	373	360	367	368	369	373	366
Female	**Mean**	56.83	0.56	0.28	14.83	29.78	0.22	0.67
	SD	13.78	0.50	0.45	11.58	19.20	0.42	0.37
	n	53	52	53	53	54	54	54
Joint (M&F)	**Mean**	45.24	0.95	0.19	16.24	31.67	0.62	0.51
	SD	11.48	0.22	0.40	13.64	11.76	0.50	0.33
	n	21	21	21	21	21	21	21
Total	**Mean**	52.34	0.91	0.67	11.27	26	0.56	0.60
	SD	12.47	0.29	0.47	10.68	14.55	0.50	0.42
	n	447	433	441	442	444	448	441

HoHH = Head of household | SD = Standard Deviation

Table S11: Summary statistics for key covariates for female HoHH by marital status

Female HoHH by Marital Status		HoHH Age	HoHH Literacy	Min. to Basic Health Service	Min to Adv. Health Service	Able to Afford Pro. Health Care	TVs by HH pop.
Married	**Mean**	55.31	0.31	16.21	27.38	0.17	0.62
	SD	14.10	0.47	10.65	12.85	0.38	0.38
	n	29	29	28	29	29	29
Widowed	**Mean**	59.95	0.27	13.82	32.45	0.32	0.79
	SD	12.32	0.46	13.05	26.26	0.48	0.31
	n	21	22	22	22	22	22

HoHH = Head of household | SD = Standard Deviation
Two missing observations for female HoHH
Marital status for one Female HoHH = "single"

Table S12: HWT method and marital status for female headed households

	Married	Single	Widowed	Total
Boil: Electric Kettle	9	0	3	12
Boil: Pot	11	0	9	20
Bottled Water	7	1	9	17
Untreated Water	0	0	1	1
Total	27	1	22	50

Table S13: Summary statistics for covariates for female HoHH by marital status if boil with pots

Female HoHH by Marital Status if Boil with pots		HoHH Age	HoHH Literacy	Min. to Basic Health Service	Min to Adv. Health Service	Able to Afford Pro. Health Care	TVs by HH pop.
Married	**Mean**	58.73	0.36	17.50	23.82	0.36	0.58
	SD	11.85	0.50	11.13	12.88	0.50	0.22
	n	11	11	10	11	11	11
Widowed	**Mean**	65	0	15.44	45.56	0.33	0.81
	SD	8.65	0	14	30.87	0.50	0.28
	n	9	9	9	9	9	9

HoHH = Head of household | SD = Standard Deviation
Two missing observations for female HoHH
Marital status for one Female HoHH = "single"

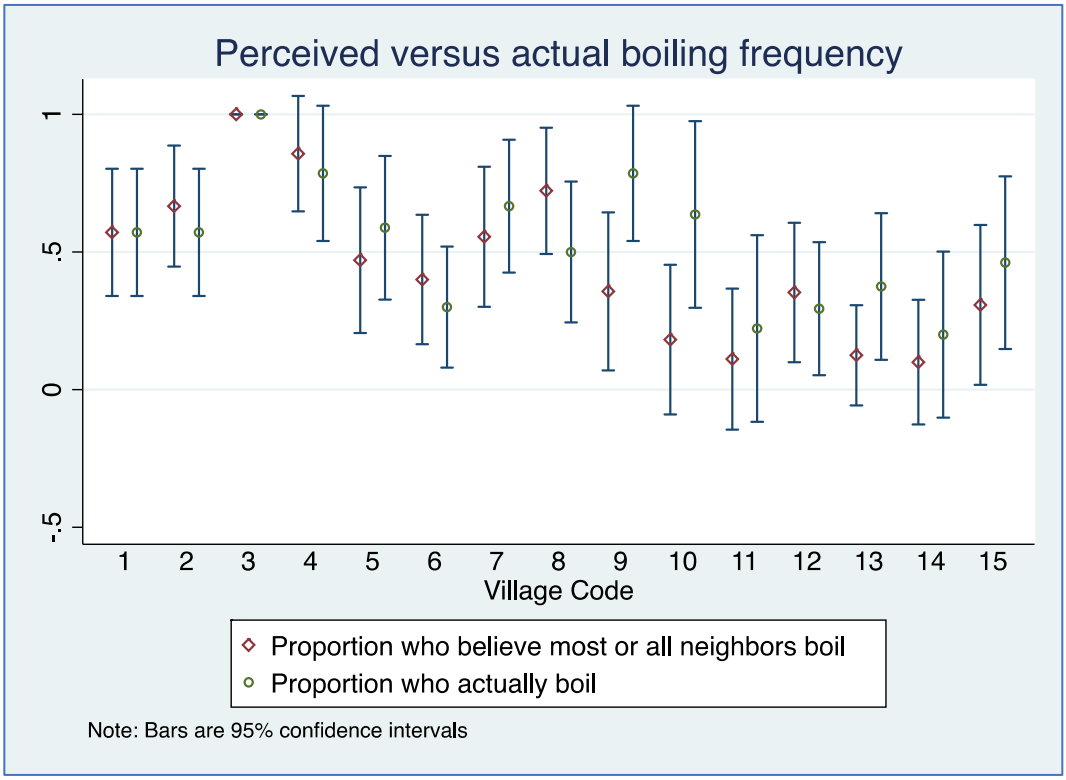

Figure S8: Perceived versus reported boiling prevalence by village

When comparing households that boil their water to those who drink untreated water or those that drink bottled water, if the household believes "most" or "all" of their relatives boil their water, they are 1.56 times more likely to boil than drink untreated water (RR=1.56, CI=1.31-1.85, p<0.0001) and 1.36 times more likely to boil than drink bottled water (RR=1.36, CI=1.09-1.69, P=0.0039). Overall, households who boil with pots had the highest level of belief that most or all of their relatives and neighbors also boil – these proportions and their respective confidence intervals are shown in Figure S6.

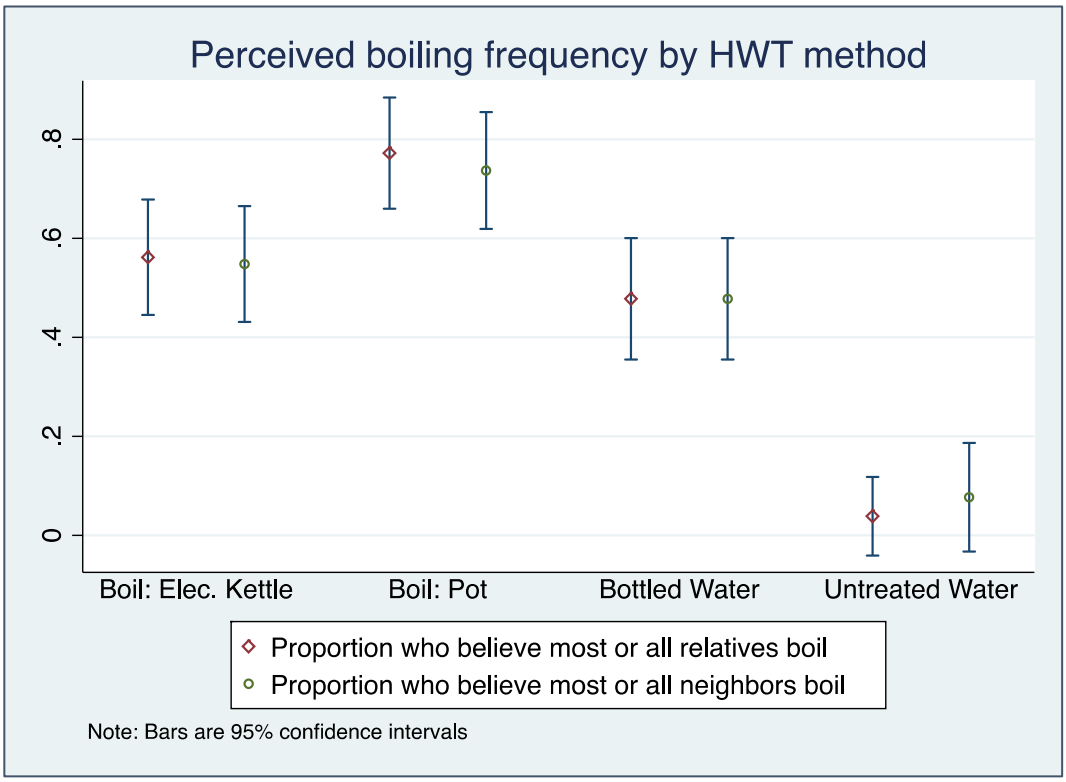

Figure S9: Perceptions of boiling prevalence among relatives and neighbors by HWT method

Additional information related to data presented in Table 3

Because we were unsure about reliability of the village-level income data, we used a number of proxies for household wealth and income in our models. In Table 3, mean per capita TV ownership is significantly lower for households in lower income villages compared to those in middle and upper income villages (0.53 vs. middle-upper combined mean of 0.63: two-sided t-test with unequal variance, $p=0.0198$), and, although the mean for the middle and upper groups are essentially the same, the SD is smaller for the high-income villages (indicating less variation in rates of TV ownership, which would be expected). Similarly, the proportion of households who report being able to afford professional healthcare (if needed) increases significantly from the lower income group to the upper, where close to 90% of households in upper income villages reported being able to afford professional healthcare. Similarly, there is a clear trend (reflected in the models) such that access to social services such as basic healthcare is worse in lower income villages than in upper income villages. It takes significantly longer to reach a health clinic in the eight villages with reported incomes below RMB 5,100 (mean=13.8 minutes, SD=12.8, n=238) than it takes in the seven villages with incomes above RMB 5,100 (mean=8.5 minutes, SD=6.6, n=205) (two-sided t-test with unequal variance, $p<0.0001$). Taken together, this suggests that, while the resolution of reported village income is crude, overall the data appear to provide an accurate indication of household incomes.

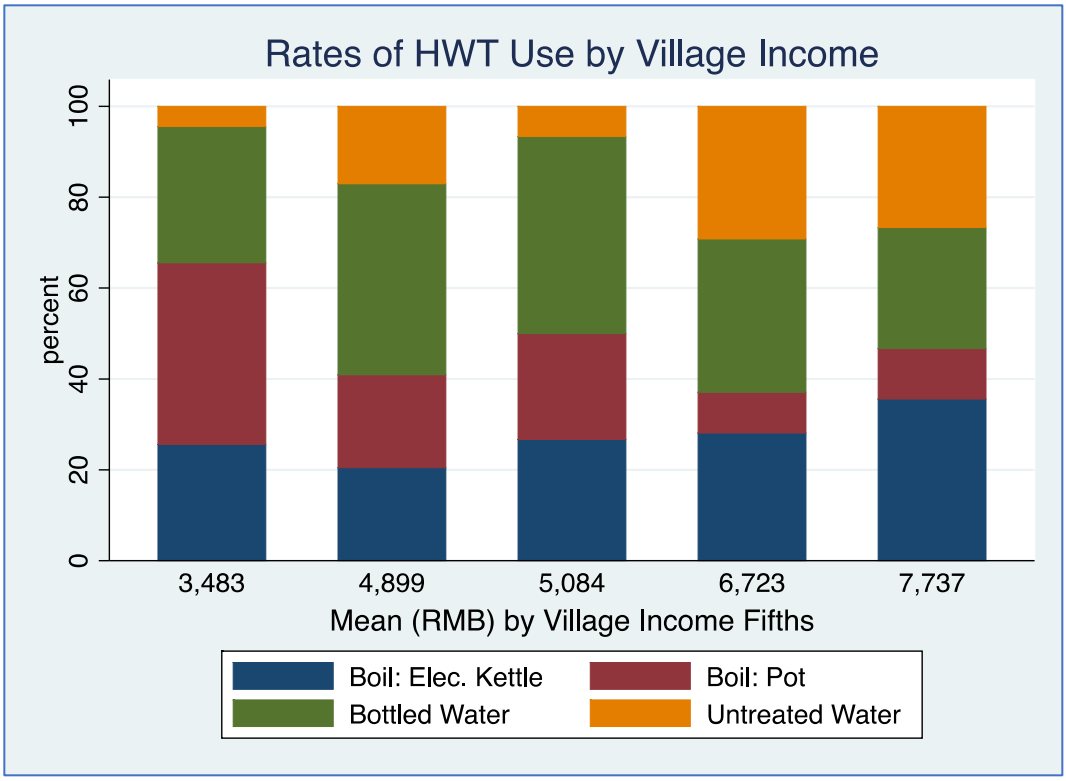

Figure S10: HWT method prevalence by village income fifths, including untreated

As can be seen in Figure S10, in higher income villages, a larger than expected proportion of household did not treat their water; through discussions with county-level staff, we learned that the spring water in these villages is believed to be of high quality (though our analyses showed otherwise) and, therefore, many households choose not to boil it.

Table S14: Bottled water costs by village income fifths (in RMB)

Mean Village Income	**Bottled Water Costs**					
	Minimum	**Median**	**Maximum**	**Mean**	**SD**	**Villages in Group**
3,483	8.88	8.96	10.90	9.58	0.94	3
4,899	4.15	7.82	9.55	7.17	2.27	3
5,084	7.84	8.51	9.24	8.53	0.57	3
6,723	6.05	6.45	7.21	6.57	0.49	3
7,737	5.03	6.05	7.40	6.16	0.98	3
Total	**4.15**	**7.82**	**10.90**	**7.60**	**1.77**	**15**

8. Discussion: Key factors associated with HWT and bottled water use

Table S15: Frequency of HWT use by head of household literacy, age, and village-income level

Village-level income	HWT method (excludes untreated)	Head of the household is:			
		Illiterate		Literate	
		Younger	Older	Younger	Older
Lower Income (n=130)	Boil: Elec. Kettle	6.9	8.5	9.2	6.2
	Boil: Pot	5.4	16.2	4.6	6.9
	Bottled Water	6.2	8.5	15.4	6.2
Medium & Upper Income (n=234)	Boil: Elec. Kettle	3.0	5.6	13.7	12.0
	Boil: Pot	3.8	5.1	4.7	6.8
	Bottled Water	4.3	4.3	24.4	12.4

Cell values are the percentage of households within each income group
Low Income = RMB 2,984 - 4,868 | Middle & High Income = RMB 5,000 - 8,526
Head of the household's age: Younger = 23-52 | Older = 53-80

Table S16: Frequency of HWT use by head of household gender, age, and income level

Village-level income	HWT method (excludes untreated)	Head of the household is:			
		Female or joint M-F		Male	
		Younger	Older	Younger	Older
Lower Income (n=132)	Boil: Elec. Kettle	3.0	3.0	13.6	11.4
	Boil: Pot	6.1	9.8	5.3	12.1
	Bottled Water	6.8	3.0	15.2	10.6
Medium & Upper Income (n=237)	Boil: Elec. Kettle	0.4	2.1	16.0	15.6
	Boil: Pot	0.4	2.5	8.0	9.3
	Bottled Water	3.0	2.1	26.2	14.3

Cell values are the percentage of households within each income group
Low Income = RMB 2,984 - 4,868 | Middle & High Income = RMB 5,000 - 8,526
Head of the household's age: Younger = 23-52 | Older = 53-80

Table S17: Reasons why household purchases bottled water: Population proportions

	County A (n=82)	County B (n=70)	Total (n=152)
Bottled water is convenient	0.537	0.386	0.462
Bottled water is safe	0.122	0.300	0.210
Bottled water is affordable	0.159	0.043	0.101
Bottled water tastes good	0.061	0.143	0.101
Because many people drink bottled water	0.049	0.100	0.074
Tap water is poor quality	0.061	0.014	0.038
Other	0.012	0.014	0.013
Totals	**1**	**1**	**1**

Cells are population proportion estimates (using sampling weights)

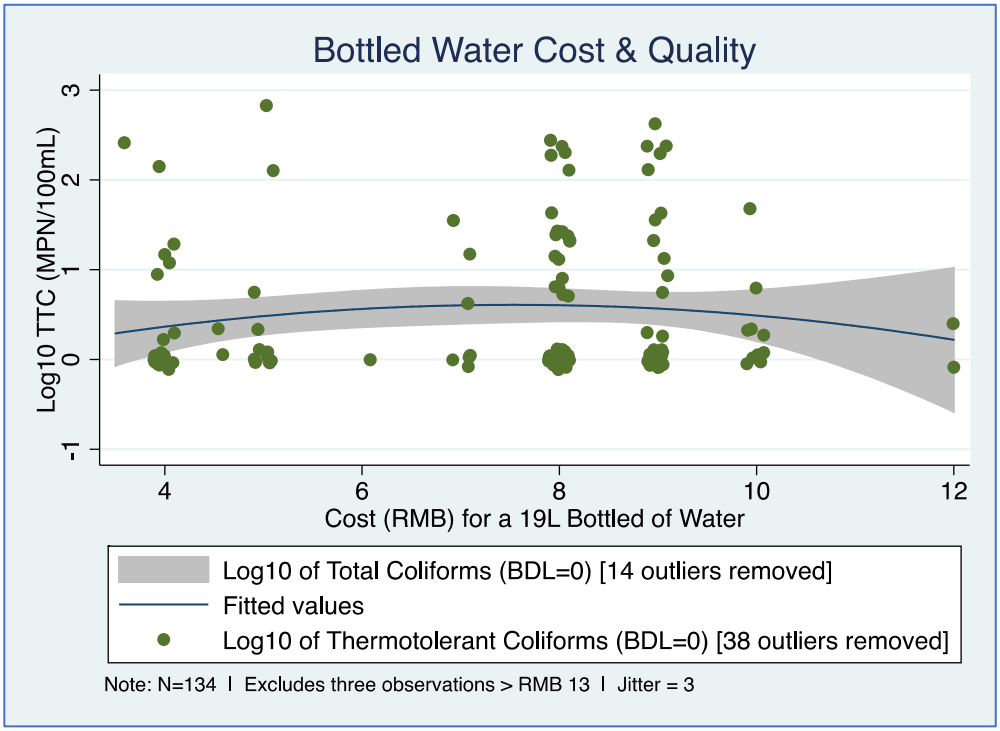

Figure S11: Bottled water cost and TTC concentrations

Table S18: Reasons why household does not purchase bottled water: Population proportions

(Survey Q104)	Boil (all) (n=206)	Boil: Electric Kettle (n=115)	Boil: Pot (n=91)	Untreated (n=70)	Total (n=276)
Too expensive	0.389	0.323	0.475	0.335	0.375
Not easy to get/purchase	0.168	0.150	0.192	0.072	0.143
Not safe	0.151	0.155	0.146	0.104	0.139
Prefer spring water	0.163	0.242	0.060	0.237	0.182*
Don't know	0.051	0.057	0.044	0.093	0.062
Don't like or tastes bad	0.021	0.028	0.012	0.044	0.027
Not needed	0.026	0.009	0.048	0.072	0.038
Other	0.030	0.036	0.023	0.043	0.033
Totals	**1**	**1**	**1**	**1**	**1**

Note: Cells are population proportion estimates (using sampling weights)

*****For the ~18% of households who reported that they did not purchase bottled water because they had access to spring water which was of high quality: these households tended to be in higher income villages, however a stratified analysis of water quality based on source did not suggest that their water was of higher relative quality.