

Using Entertainment-Education to Promote Safer Sexual Behavior in Young Adults in Nigeria

Victor Hugo Orozco-Olvera

Department of Social Policy and Intervention

Wolfson College

Trinity 2019

Thesis submitted in partial fulfillment of the degree of Doctor of Philosophy
in Social Intervention by stand-alone papers at the University of Oxford.

Word count: Approximately 59,000 excluding references and appendices

VOLUME I

***“Reason Is and Ought Only
to Be the Slave of the
Passions”***

David Hume

Publications and presentations arising from this thesis

Published

Orozco, V., Shen, F., and Cluver, L. (2019). The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017. PLOS ONE. doi.org/10.1371/journal.pone.0209969
(Paper 1)

Banerjee, A., La Ferrara, E., **Orozco, V.** (2019). Entertainment, Education and Attitudes Towards Domestic Violence. Papers and Proceedings. American Economic Association May 2019 edition

Under review or to be submitted

Banerjee, A., La Ferrara, E., **Orozco, V.** The Entertaining Way to Behavioral Change: Fighting HIV with MTV

Orozco, V. Does an Educational Television Series Work for the Population Where It Matters Most? A Post-hoc Moderators Analysis of Youth Engaging in High-Risk Sexual Behaviors in Nigeria. **(Paper 2)**

Orozco, V. and Cluver, L. How did an Entertainment-Education intervention promote HIV testing? A Post-hoc Mediation Analysis of the MTV Shuga Television series. **(Paper 3)**

Academic presentations

Campbell Collaboration, "What Works Global Summit" (London, UK), 2016
Instituto Tecnológico Autónomo de México (ITAM)- Research and Policy Center (Mexico City, Mexico), 2016

Johns Hopkins University, School of Advanced International Studies, Research seminar, (Washington DC, US), 2016

Middlesex University, Economics Department research seminar (London, UK), 2017

International Communication Association Conference (San Diego, US), 2017
Oxford University, Department of Social Policy and Intervention, CEBI seminar (Oxford UK), 2017 and 2019
Zurich University, Department of Economics, IV Annual Conference: What Works? Social and Behavioral Approaches to Boost Children's Well-being (Zurich, Switzerland), 2019
Centre for Research on Impact Evaluation- European Commission (Ispra, Italy), 2019

Policy presentations

Department for International Development (DFID) Behavioral Change Unit seminar (London, UK), 2017
Federal Ministry of Health of Nigeria, World Bank Impact Evaluation Workshop (Abuja and Lagos, Nigeria) 2016 and 2017
World Bank Headquarters (Washington DC, US)

- a. Development Impact Evaluation Unit-Behavior Change Community of Practice, "Preliminary Results of the MTV Shuga evaluation", 2016
- b. Independent Evaluation Group, "Innovative Research in Fragile States", 2016
- c. Poverty Global Practice, "Methodological Innovations in Evaluation Research", 2017
- d. World Bank Behavioral Science for Development: A Storytelling event, organized by the Bank's behavioral economics or nudge unit.

Gates Grand Challenges Annual Summit (Washington DC, US) 2017
World Bank and International Monetary Fund Spring Meetings (Washington DC, 2017)
Elevate Entertainment-Education Conference (Mumbai, 2018)
World Bank Narrating Behavior Change Workshop (New Delhi, 2018)
World Bank Econothon 24-hour online event (Washington DC, 2019)

THESIS ABSTRACT

Background: Systematic reviews show that HIV prevention campaigns are often ineffective in reducing risky sexual behaviors. Promoting safer sexual practices among high-risk individuals is key in curbing the HIV epidemic. Entertainment-education is the use of mass media programming with educational and development objectives. MTV Shuga is an educational television series aimed at promoting safer sexual practices among Africa youth. Using data from a randomized control trial in Nigeria, this thesis studies the effectiveness of MTV Shuga on promoting safer sexual practices among the high-risk youth from the general urban population; and investigates the mediating effects of psychosocial and immersion mediators on the program's impacts on increased HIV testing.

Methods: This thesis is comprised of three studies. The first is a meta-analysis of entertainment-education narratives conducted in field settings in developed and developing countries (n=10 studies; 23,476 individuals). Effect sizes were meta-analyzed using random-effects models. The moderators paper is the second study of this thesis. This paper studied the average treatment effects of MTV Shuga on high-risk youth (n=2,073), including average treatment effects by gender (sub-group analysis); and the moderating effects of religion and television-watching habits for this high-risk population (effect-modification analysis). Data was analyzed from 18-25-year-olds from the urban general population that at baseline reported at least one of the following HIV-risk factors: had unprotected sex in their last sexual encounter, engaged in sexual concurrency, had a sexual partner older/younger by ten years, or reported having symptoms of a STI. The moderators analysis is based on intent-to-treat estimates.

The mediation analysis is the third paper of this thesis. It aimed to identify the mechanisms through which MTV Shuga impacted one of the program's objectives, the uptake of HIV testing. The study sample is restricted to individuals that were sexually active at baseline and for whom the Shuga trial had data for HIV testing and all tested mediators at follow-up (n=1,894). The analysis is based on the mediation approaches of Hicks and Tingly (2011) and Preacher and Hayes (2008). Given methodological limitations and the post-hoc nature of the analysis, the results of the thesis should be read as hypothesis-generating evidence.

Results: In the first study, the meta-analysis found suggestive evidence that entertainment-education narratives had small but statistically significant effects for three out of four high-risk sexual behaviors. This communication approach reduced the number of sexual partners, reduced unprotected sex rates, and increased testing and management for STIs. However, entertainment-education was not effective in reducing inter-generational sex. It had medium-size effects on knowledge outcomes, where a time-decay relationship is observed. This study found no effects on HIV attitudes.

The second study provides suggestive evidence that MTV Shuga increased HIV testing among high-risk individuals, and decreased the average number of current sexual partners. It increased knowledge about HIV transmission and reduced stigma and negative attitudes towards HIV-positive people. However, this paper did not find effects on chlamydia infections, condom use, and inter-generational sex. This study provides indicative evidence of consistency of program effects across different populations: Program effects for high-risk youth were similar to results reported for the average study

participant. Program effectiveness among high-risk youth did not differ by participant's religion or television-watching habits.

In the third study, the mediation analysis found that impacts on a self-reported HIV testing measure (eight-month follow-up survey) were mediated by knowledge about HIV transmission, attitudes towards HIV+ people and program transportation. The latter had the strongest influence, mediating a quarter of the total effect. For an objective HIV testing measure collected two months after the follow-up survey, program transportation remained the most influential variable though it was no longer statistically significant.

Conclusion: This thesis shows that mass media entertainment may be an effective tool in promoting safer sexual practices among youth. Although it generally found small behavioral effects, the economies of scales of entertainment-education over face-to-face interventions suggest its potential to be a cost-effective tool above an audience threshold. The thesis provide indicative evidence about the “transportability” of entertainment-education narratives. Partly mediated by program transportation, MTV Shuga's effectiveness on psychosocial and behavioral outcomes was similar for different youth populations in Nigeria. Given the limited effectiveness of traditional HIV prevention campaigns, research is needed to understand how best these campaigns could be complemented with entertainment-education modules.

THESIS STRUCTURE

This thesis is organized in six chapters. Chapter One serves as an introduction to the thesis, discussing the theoretical framework of entertainment-education narratives, the current knowledge gaps of the literature, and the research questions the thesis intends to

address. Chapter Two is the Methods Chapter, describing the methods used in all three DPhil papers. Chapters three, four and five present the meta-analysis (Orozco, Shen and Cluver (2018), and the MTV Shuga moderators (Orozco 2019) and mediators (Orozco and Cluver 2019) papers respectively. The three papers are formatted for public health journals. For publication objectives, the papers summarize the methods described in the Methods Chapter. Despite the best efforts of the doctoral candidate to reduce repetition across the thesis chapters, some repetition persists. Chapter Six summarizes findings of the three papers, concluding with recommendations for policy and future research.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of numerous individuals and institutions. The generous support from the Bill and Melinda Gates Foundation and the World Bank's i2i program made this study possible. The scholarship from Mexico's *Consejo Nacional de Ciencia y Tecnología* is greatly appreciated.

I wish to thank Georgia Arnold and the MTV Staying Alive Foundation for supporting this study from the very beginning; Segun Oguntinyinbo and his professional team of enumerators at Hanovia Medical Limited that helped carry out thousands of surveys; Olusegun Fasina and the College of Medicine-University of Ibadan for conducting the temporary clinics; and Maada Gombu and Breaking Barriers in Development for conducting the community screenings to study the Shuga program. I want to thank Edwin Ikhuoria, Dante Donati and Laura Costica and the many other field coordinators and research assistants that helped supervise the intervention and the data collection. I want to thank the different authors that provided their data for the meta-analysis, especially Professors Vaughan and Singhal who through distinct conversations also advised on the theoretical framework.

I wish to thank my supervisor Professor Lucie Cluver. I decided to come to Oxford because in my first meeting with her, it was clear that she would push me to do my best. I am glad she did. Through her insightful and detailed guidance, she has implanted in me a desire to excel in my research. Her dedication to rigorous and policy relevant research inspires me to pursue an academic career while closely working with policymakers. My advisor's patience for reviewing what may have been an unusual number of drafts, under very tight deadlines due to my late submissions, will always be remembered.

This thesis would not have been possible without my parents, Teodora and Carlos. Their emotional support during the dark days of my doctorate helped me to carry on. I want to thank Arianna Legovini, my manager at the World Bank, who supported my DPhil research from day one by cutting my other work obligations. Big thanks to so many colleagues and friends who patiently reminded me about basic STATA commands and who reviewed my many drafts. Many thanks to my late father-in-law Dan Salcedo whose evening babysitting allowed me to stay longer at the office running regressions and re-writing the thesis. But above all, I want to thank my wife Adriana for allowing me to dedicate so many evenings, nights and weekends to finish my thesis. Her patience and emotional support will always be cherished.

CANDIDATE'S ROLE AND CONTRIBUTION

Candidate's role in the Shuga trial

The doctoral candidate has been working as an Economist at the World Bank research group (DECRG) since 2009, where he has launched a series of randomized control trials (RCT) of HIV behavior change programs in Sub-Saharan Africa. In this capacity, he launched the Shuga trial in 2013. The candidate is one of three Principal Investigators, the other two are academics external to Oxford University. This external arrangement was due to delays in the candidate's DPhil funding and the pressure from study donors to start the trial. The candidate is an equal and corresponding co-author for the main paper of the Shuga trial (co-authorship is in alphabetical order, as its traditionally done in economics). The main outcomes paper is referred in the thesis as Banerjee, La Ferrara and Orozco (2018).

The doctoral candidate has played a leading role in the Shuga trial. Prior to working with the external academics, the candidate independently: secured four research grants that fully financed the study (two from the Bill and Melinda Gates Foundation and two from the World Bank i2i research window) and developed the evaluation design with the study clients: Nigerian National AIDS Commission and MTV Staying Alive Foundation. The candidate was responsible for hiring and supervising a broader research team of research assistants and field-based coordinators, a survey firm, a medical school, and an implementing NGO to carry out the study intervention. This team is listed in research protocols submitted to different ethical committees, including Oxford's OxTREC. The candidate conducted four fieldwork visits in 2014 and 2015, for a total of 3 months. The candidate led policy dialogue with government officials, the piloting of the survey

instruments and the training delivery for enumerators and clinicians. Present in the first two weeks of data collections, the candidate remotely supervised the collection and the cleaning of the baseline and follow-up datasets; and actively worked with the external academics and the broader research team in refining the evaluation design and instruments, as well as in the final data analysis and write-up of the main paper. This paper, which can be found in Appendix 2, is not part of the thesis as the candidate is an equal but not a lead co-author as required by University guidelines.

Candidate's role in the DPhil papers

The doctoral candidate was the lead investigator for the three papers required for obtaining the DPhil degree in Social Intervention. The candidate conducted the design, data analysis and write-up for all three papers. The candidate is the single author for the moderators paper and lead-co-author for the meta-analysis and the mediators papers. Professor Lucie Cluver, the candidate's doctoral advisor, provided guidance and detailed comments to all three DPhil papers.

Declaration of authorship

I, Victor Orozco, hereby declare that this thesis is my own work except, where otherwise stated.



Signed _____

Oxford 04/11/2019

© Victor Orozco, 2019

The thesis, parts thereof, figure and tables may not be copied or reproduced without prior written permission of the author, Victor Orozco.

TABLE OF CONTENTS

VOLUME I

CHAPTER 1. INTRODUCTION	23
Background	23
Research Objective and Research Questions	27
Paper 1: “The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017”	27
Paper 2: “Does an educational television series work for the population where it matters most? A post-hoc moderators analysis of youth engaging in high-risk sexual behaviors in Nigeria”	28
Paper 3: “How did an entertainment-education intervention promote HIV testing? A post-hoc mediation analysis of the MTV Shuga television series”	28
Theoretical Framework	29
Testing the theoretical framework from a moderation perspective	35
Testing the theoretical framework from a mediation perspective	35
Current Gaps in the Literature	37
Thesis Contributions to the Literature	40
 CHAPTER 2. METHODOLOGICAL FRAMEWORK	 48
Meta-analysis of Entertainment-Education Narratives (Paper 1)	49
Inclusion Criteria	49
Search Strategy	49
Risk of Bias Assessment and Quality of Evidence	50
Data Analysis	50
Moderators and Mediators Analyses of the MTV Shuga Trial (Paper 2 and 3)	52
Research Setting	52
Study Sample	54
The MTV Shuga Trial	57
The Intervention	57
Evaluation Design	57
Sampling and Data Collection Strategy	67
Sample Size	69
Survey Planning and Quality Controls	70

Recruitment Strategy	71
Data Collection	72
Trial Flow Diagram and Follow-up Attrition	79
Data Challenges	80
Ethical Considerations	80
Informed Consent	80
Privacy, Data Storage and Security	81
Potential Risks of Participation	82
Ethical Approval	82
Measures	83
Outcome Selection for Thesis	83
Psychosocial Variables	87
Entertainment-Education Variable	89
Behavioral Variables	89
Construction of Variables	92
Statistical Analysis	93
Technical Considerations for the Thesis Papers	93
Moderator Analysis	97
Mediators Analysis	101
Single mediator Models	103
Multiple Mediators Model	106
CHAPTER 3. Paper 1: The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017	107
Abstract	108
Introduction	110
Methods	112
Selection Criteria	113
Search Strategy	113
Risk of Bias Assessment	114
Data Analysis	114
Results	116
Results of the Search	116
Key Features of Included Studies	116

Measures of Risky Sexual Behavior	120
Effectiveness on Safer Sexual Practices	121
Meta-analysis Subgroup Analysis	123
Knowledge and Attitudes	123
Quality of the Evidence	125
Discussion	126
Supplementary Information	129

CHAPTER 4. Paper 2: Does an educational television series work for the population where it matters most? A post-hoc moderators analysis of youth engaging in high-risk sexual behaviors in Nigeria 151

Abstract	152
Introduction	154
Methods	156
Study Design	156
Study Location and Participants	157
Measures and Outcomes	159
Study Attrition	163
Statistical Analysis	164
Results	166
Randomization Check and Sample Characteristics	166
Subgroup Analysis for High-risk Youth, by Gender	169
Effect-modification Analysis of Religion and Television-watching Habits	172
Discussion	174
Conclusion	178
Paper Appendix Tables	180

CHAPTER 5. Paper 3: How did an entertainment-education intervention promote HIV testing? A post-hoc mediation analysis of the MTV Shuga television series 182

Abstract	183
Introduction	185
Methods	191
Study Design and Sample	191
Causal Mediation Framework and Statistical Analysis	192

Single Mediator Models	192
Multiple Mediators Model	193
Measures	194
Short-term and Long-term HIV Testing	194
Mediator Variables	195
Statistical Analysis	196
Results	197
Randomization Check and Sample Characteristics	197
Short-term Effects (0 to 8 months)	199
Long-term Effects (8 to 10 months)	202
Discussion	204
Conclusion	207
Paper Appendix Tables	209
CHAPTER 6. Summary and Discussion of Thesis Findings	210
Overview	210
Discussion of Findings	213
Paper 1: The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017	213
Paper 2: Does an educational television series work for the population where it matters most? A post-hoc moderators analysis of youth engaging in high-risk sexual behaviors in Nigeria	221
Paper 3: How did an entertainment-education intervention promote HIV testing? A post-hoc mediation analysis of the MTV Shuga television series	227
Strengths of the Thesis	230
Limitations of the Thesis	233
Directions for Future Research	235
Policy Implications	239
Dissemination	241
Next Steps	243
REFERENCES	247

VOLUME II

APPENDICES

Appendix 1. Registration Form of Systematic Review	261
 Appendix 2. (Main Paper of the Shuga Trial). The entertaining way to behavioral change: Fighting HIV with MTV	
Introduction	272
A Simple Model of Learning and Conformity	275
Background and Experimental Design	280
Empirical Strategy and Data	283
Results	289
Average Treatment Effects	289
Social Effects	289
Conclusions	295
Study Tables	297
 Appendix 3. Script and Main HIV/AIDS Messages of MTV Shuga (season 3)	
316	
 Appendix 4. TIDieR Checklist (Template for Intervention Description and Replication)	
318	
 Appendix 5. Consent Forms for Surveys and STI Biomarkers Used in the Shuga Trial	
322	
 Appendix 6. Ethics Approval Letters (Nigerian Government Ethics Committee, Massachusetts Institute of Technology and Bocconi University)	
326	
 Appendix 7. Ethics Application Submitted to Oxford Ethics Committee (OxTREC)	
330	
 Appendix 8. Ethics Approval Letter (OxTREC)	
353	
 Appendix 9. Survey Instruments	
9A. Baseline	355
9B. Eight-month Follow-Up	416

Appendix 10. Interviewer’s Instruction Manual

Introduction	486
General Guidelines, Informed Consent and Introductions	491
Electronic Surveys - Handling the Google Nexus 7 Tablet	494
Survey Instrument Structure	502
Special Instructions for Each Module	503

LIST OF TABLES AND FIGURES

INTRODUCTION AND METHODOLOGICAL FRAMEWORK

Table 1. Comparison Shuga sample with urban population-level survey	57
Table 2. Evaluation design of the MTV Shuga trial	64
Table 3. Timeline for the average study participant	75
Figure 1. Theory of change of entertainment-education narratives	34
Figure 2a. Logic Model of the Shuga trial	45
Figure 2b. Conceptual framework for studying variation in program effects, treatment contrasts and implementation	47
Figure 3: Study states in southwest Nigeria	54
Box 1. Example of Smart Graph used for Treatment 2	65
Figure 5. Shuga flow diagram	79
Figure 6. Testing constructs of the theory of change of entertainment-education	102

Paper 1: “The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017”

Table 1. Descriptive Characteristics of 10 included studies of entertainment-education narratives on youth sexual health	119
Figure 1. Flow Diagram of study selection process, 1985-2017	118
Figure 2. Effects on four sexual behaviors	122
Figure 3. Effects on knowledge	124
Figure 4. Effects on attitude	125

Supplementary information – Meta-analysis

Table S1. PRISMA Checklist	129
Table S2. Intervention and Reason for Exclusion	132
Table S3. Characteristics of Included Studies and Support for Risk of Bias Assessment	135
Table S4. Effects on Individual Behavior Outcomes	141
Table S5. Summary of Findings for the Main Comparison	142
List S1. MEDLINE search strategy	144
List S2. List of 36 studies excluded from the meta-analysis	146
Figure S1. Funnel Plot for Primary Analysis (n=8)	149
Figure S2. Meta-regression Effect Size vs Timing of the Follow Up	150

Paper 2: “Does an Educational Television Series Work for the Population Where It Matters Most? A Post-hoc Moderators Analysis of Youth Engaging in High-Risk Sexual Behaviors in Nigeria”

Table 1. Socio-demographic and HIV outcomes, Pre-intervention characteristics	168
Table 2. Average Treatment Effects for high-risk youth, gender breakdown	170
Table 3. Moderating Effects of Religion and Television-Consumption Levels	173
Table A1. Comparison Shuga Sample with Population-Level Survey	180
Table A2. Pre-intervention comparison of Sample High-risk pool versus Low-risk pool	181
Table A3. Attrition Rates for Follow-Up and temporary Clinics	181

Paper 3: “How did an Entertainment-Education intervention promote HIV testing? A Post-hoc Mediation Analysis of the MTV Shuga Television series”

Table 1. Socio-demographic and HIV outcomes, Pre-intervention characteristics	198
Table 2. Mediating effects for short-term outcome	201
Table 3. Mediating effects for long-term outcome	203
Table A1. HIV Testing and Potential Mediators	209
Figure 1. Effect timeline on increased HIV testing	188
Figure 2. Testing constructs of the theory of change of entertainment-education	190

KEY TERMS

ACME	Average causal mediation effect
ANCOVA	Analysis of covariance
ARV	Anti-retroviral
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
df	Degrees of freedom
GPS	Global Positioning System
GRADE	Grading of Recommendations Assessment, Development, and Evaluation criteria
HIV	Human immunodeficiency virus
ICC	Intra-cluster correlation
ITT	Intention-to-treat
N	Sample size
NDHS	Nigeria Demographic and Health Survey
NGO	Non-governmental organization
OxTREC	Oxford Tropical Research Ethics Committee
p.p.	Percentage points
pca	Principal component analysis
PEPFAR	President's Emergency Plan For AIDS Relief
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomized Control Trial
SD	Standard deviations
SMD	Standardized mean difference
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
STI	Sexually transmitted infections
TIDieR	Template for Intervention Description and Replication Checklist
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
WHO	World Health Organization

CHAPTER 1. INTRODUCTION

This thesis studies the effectiveness of entertainment-education narratives on youth risky sexual behaviors. It is based on a global meta-analysis and an effectiveness trial conducted in southwest Nigeria. This chapter serves as an introduction to the thesis, discussing the study's background, the theoretical framework of entertainment-education narratives, the current knowledge gaps of the literature and the research questions the thesis aims to address.

Background

In 2017, an estimated 1.2 million people in sub-Saharan Africa became infected with HIV. More than one-third of these infections occurred in youth between the ages of 15 and 24 (UNAIDS 2018). With an estimated population of 160 million, Nigeria is second to South Africa in the number of people living with HIV/AIDS worldwide, representing 3.2 million or 9 percent of the global burden of the disease (Nigeria Global AIDS Response Report 2017). Important challenges remain for Nigerian youth. According to Nigeria's 2013 Demographic and Health Survey (2013 NDHS), only one-third of 15-24-year-old youth had comprehensive knowledge about AIDS; and one in nine sexually active youth knew their HIV status. Unprotected heterosexual sex represents 80 percent of new cases of HIV in Nigeria (National Agency for the Control of AIDS. 2015. Country Progress Report); and the leading infection source among youth in Sub-Saharan Africa and worldwide (UNAIDS 2016; UNICEF 2017).

Systematic reviews of HIV behavior change programs in low and middle-income countries (Fonner et al. 2014, McCoy et al. 2010), including Sub-Saharan Africa (Harrison et al. 2010;

Michielsen et al. 2010, Mavedzenge et al. 2011), show these programs generally improved knowledge and awareness, but had limited impacts on reducing risky sexual behaviors. In a review of systematic reviews, Krishnaratne et al. (2016) looks at the effectiveness of a series of HIV prevention interventions. While this review found strong evidence for a series of medical and supply-side interventions, it concluded that results for HIV behavior change interventions were overall “disappointing”¹. However, this and other reviews note low quality of studies, which may partly explain the overall flat results. Systematic reviews for interventions that target individuals at high risk of HIV infection find mixed evidence for this subpopulation. However, most reviews are based on resource-intensive interventions conducted in the US (Abad et al. 2015; Herbst et al. 2007), which partly limits their relevance for low-resource settings.

The limited effectiveness of HIV information-based interventions has led researchers to test more resource-intensive interventions, including biomedical and cash transfers. Research in the last decade found effectiveness evidence of voluntary male medical circumcision (Auvert et al. 2005; Bailey et al. 2007; Gray et al. 2007) and pre-exposure prophylaxis for preventing HIV transmission (Cohen et al. 2011). However, the issues of program uptake and sustained behavior change remain important challenges for biomedical prevention strategies. Recent HIV research suggests that cash transfers can be effective tools in addressing both issues. Through a prospective longitudinal study, Cluver et al. (2016) found that the combination of a series of social protection services and cash transfers reduced risky sexual behaviors among teenagers in South Africa. Experimental field studies have shown that financial incentives can increase HIV testing uptake in

¹ Behavioral economics research shows this challenge is not unique to HIV, as promoting behavior change is a challenge observed in other development sectors (O’Donoghue and Rabin 2001; World Bank 2015).

Malawi (Kohler and Thornton 2011), reduce unprotected sex and sexually transmitted infections (STIs) in Tanzania (de Walque et al. 2012), and increase adherence to HIV treatment in the US (El-Sadr et al. 2017).

However, fiscal restrictions in low-income countries make some of these resource-intensive programs difficult to implement at a larger scale. In countries with generalized HIV epidemics (i.e., where HIV prevalence rates among the general population exceed 1 percent), the policy and fiscal advantages of effective mass media interventions are clearer because most sexually active people are at an elevated risk of acquiring HIV. UNAIDS (2005 and 2007) recommends targeting “all segments of society” in this context. Thus, *effective* and *low-cost* HIV information interventions that reach substantial segments of a population are likely to remain for the foreseeable future as the main tools for containing the epidemic in low-income countries (Coates et al. 2008).

Delivering HIV information to motivate safer sexual practices

In view of the rapid spread of mass media in developing countries (Moreau 2011; World Bank 2016), the use of media entertainment to deliver HIV behavior change messages has been suggested as a low-cost alternative (Singhal 2003). Government and development partners have extensively used mass media with sexual health objectives (Brown and Keller 2000; Singhal and Rogers 1999).

Entertainment-education is the “process of purposely designing and implementing a media message to both entertain and educate, in order to increase audience members’ knowledge about an issue, create favorable attitudes, shift social norms, and change the behavior of individuals and communities” (Singhal and Rogers 2004). Youth is a common target group of entertainment-education interventions (Gibbons and Gerrard 1994).

Entertainment-education that use a narrative format² (e.g., dramas, sitcoms), in contrast to those based on information-only formats (e.g., public service announcements, flyers), has received increased attention in policy circles. For example, entertainment-education is discussed as a promising behavioral tool in the World Bank's 2015 World Development Report "Mind, Society and Behavior". As discussed below in the Theoretical Framework section of this Chapter, advocates of entertainment-education narratives argue these are more effective in promoting behavior change due to (i) the "universal" appeal of stories (Sabido 2002); (ii) the power of entertainment media to emotionally charge educational messages (Cohen 2001; Murphy et al. 2011) and (iii) its ability to update audiences' perceptions of social norms (Singhal and Rogers 2004).

It is a longstanding debate on how exposure to entertainment media may affect sexual behaviors (Lowry and Shidler 1993; Page et al. 2000). While observational evidence suggests a relation (Haferkamp 1999; Walsh-Childers and Brown 1993; Anita et al. 2008), determining whether such exposure influences beliefs and behaviors with observational data is empirically difficult due to sample-selection issues (Heckman 1979; Kearny and Levin 2014). In other words, comparing the beliefs and practices of audiences from other community members that decided not to watch a program is likely to be confounded by other factors. In the last decade, a series of experimental field trials have emerged that better address these selection issues (see DellaVigna and La Ferrara 2015 for a review). Although sexual health is among the most common applications of mass media interventions (Pappas-DeLuca et al. 2008; Singhal 2001), the evidence base remains

² Kreuter et al. (2007) defines narratives as "a representation of connected events and characters that has an identifiable structure, is bounded in space and time, and contains implicit or explicit messages about the topic being addressed".

mostly qualitative (Moyer-Guse and Nabi 2011; Kearney and Levine 2015). Previous systematic reviews highlight the lack of high-quality evidence in the field of mass media and HIV (Bertrand et al. 2006; Lacroix et al. 2014; Shen and Han 2014).

Research Objective and Research Questions

The aim of this thesis is to increase the evidence-base of entertainment-education narratives with youth sexual health objectives. Paper 1 seeks to further our understanding of their effectiveness; through a meta-analysis that included the best available evidence from both developed and developing countries were included. Paper 2 studies the impacts of entertainment-education on high-risk individuals from the broader youth population in Nigeria. Paper 3 investigates the mechanisms or mediators through which entertainment-education may work in increasing HIV testing. Each paper includes an abstract that summarizes the study's background, methods, results and conclusions.

The research questions of each paper follow:

Paper 1: “The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017” (Orozco, Shen and Cluver).

PLOS ONE, February 12, 2019. <https://doi.org/10.1371/journal.pone.0209969>

- What is the impact of entertainment-education narratives on youth adopting safer sexual behaviors?
- What is their impact on youth's knowledge and attitudes?
- Are meta-analyzed effects moderated by communication medium, gender, country income level, country HIV/AIDS epidemiological level, and the type and

the intensity of the intervention? Are effect sizes moderated by study design characteristics?

Paper 2: “Does an Educational Television Series Work for the Population Where It Matters Most? A Post-hoc Moderators Analysis of Youth Engaging in High-Risk Sexual Behaviors in Nigeria” (Orozco)

- Does the MTV Shuga television drama improve HIV knowledge, attitudes towards HIV-positive people, increase HIV testing and reduce risky sexual behaviors for individuals engaging in high-risk sexual behaviors?
- What are its impacts for high-risk females and high-risk males?
- Did the religion and the television-watching habits moderate program impacts for this population?

Paper 3: “How did an Entertainment-Education intervention promote HIV testing? A Post-hoc Mediation Analysis of the MTV Shuga Television series” (Orozco and Cluver)

- Are program effects on HIV testing mediated by
 - Knowledge about HIV transmission and the benefits of HIV testing?
 - Stigma and attitudes towards HIV-positive people?
 - Transportation or immersion in the program?
- Which of the three potential mediators had the strongest mediating influence on increased HIV testing?

Theoretical Framework to the Thesis

Storytelling has commonly been used in transmitting prescriptive and normative information in most cultures for thousands of years (Sabido 2002). However, didactic forms of communication have gain ground over narratives in conveying modern health messages (Murphy et al. 2015). The theory of change most used in entertainment-education research is based on three theories: the theory of planned behavior (Ajzen 1991 and 2014), social cognitive theory with adaptations for mass media communication (Bandura 2002, 2004) and the extended elaboration likelihood theory (Cacioppo and Petty 1984; Slater and Rouner 2002). The first two are established psychology theories that explain how information-provision, social incentives and role-modeling can encourage individuals to adopt new behaviors. While the former focuses on an individual's assessments of the consequences for undertaking a certain behavior, social cognitive theory focuses on the cognitive process of observing and imitating models. Both theories are central to rational human behavior models, commonly used as theoretical frameworks of HIV information interventions (Albarracín 2001). The extended elaboration likelihood theory is greatly restricted to entertainment-education narratives. It discusses the effects that emotional reactions may have in making audiences more receptive to embedded messages. The extended elaboration likelihood theory uses psychosocial constructs, such as self-efficacy, central to the first two psychology theories.

Psychosocial Constructs

At the most basic level, all information-based interventions may help update audiences' *behavioral beliefs* of the negative consequences of engaging in risky sexual behaviors. For example, MTV Shuga aims to correct the misperception among young women that older

sexual partners are safer in terms of HIV-risk than younger partners; with the goal for women to either shift to younger partners or to increase condom use with older ones (Anabwani et al. 2016)³. The theory of planned behavior (Ajzen 1991 and 2014) asserts that whereas impacts on knowledge are required for updating behavioral beliefs (objective consequences of an action), impacts on *attitudes* or preferences for that action are required to change behavioral intentions. Attitudes are greatly determined by how retrievable the belief is from the individual's mind (Brady et al. 2008; Dale 1969)⁴.

Normative beliefs, an individuals' beliefs about the extent to which significant others (Bandura 2004) think they should perform or not perform an action, are potential motivators for adopting safer sexual behaviors (Fishbein and Guinan 1996). Normative beliefs can be an important determinant in behavioral intentions given the molding effect of internal group pressure toward uniformity (Crandall 1988; Bursztyn and Jensen 2017; Bursztyn, Fujiwara and Pallais 2017; Perez-Truglia 2017); and concerns of group sanctions for deviating from the group standard (Brown and Abrams 2003, Sussman et al. 1988, Munshi and Myaux (2005). For social norms to mediate program impacts, interventions need to target the reference group relevant for a given subpopulation and intended behavior (Fujiwara and Pallais 2017; Perez-Truglia 2017).

Individuals will be more likely to enact a behavior when they feel they can successfully do so (perceived behavioral control); which in turn is driven by feelings of self-efficacy and

³ Recent research in Botswana (Anabwani et al. 2016) finds that young women have misconceptions that young men have higher HIV prevalence rates than older men. As a result, they are less likely to request condom use with older than with younger partners.

⁴ Research has long established that images are more likely to be remembered than words (Brady et al. 2008), improving the retrievability of a belief. For example, Dale (1969) found that after three days of receiving a lecture, an individual retained only 10 to 20 percent of written or spoken information but almost 65 percent of visual information.

controllability (Bandura 1977). While self-efficacy assesses one's competence to perform an action successfully, controllability assesses the external factors and one's ability to overcome them.

A program seen by an important share of a population may potentially update audiences' perceptions of social norms (Singhal and Rogers 2004). In other words, a popular program may signal to audiences that other community members, especially "significant others" (i.e., reference group relevant for the intended behavior), are more accepting of presently stigmatized behaviors or subgroups (e.g., HIV-positive people). This in turn could potentially improve the *normative beliefs* of audiences toward the stigmatized behavior or subgroup. According to the theory of planned behavior (Ajzen 1991 and 2014), normative beliefs are an important psychosocial construct behind intended and realized behaviors, including sexual behaviors (Fishbein and Guinan 1996). For the case of mass media programs, improvements in normative beliefs can theoretically promote a sense of collective efficacy (Bandura 1997), a group's shared belief in its conjoint ability to organize and execute an action. Improvements in collective efficacy could potentially reduce audiences' perceptions of external barriers and improve a sense of controllability (Papa 2000; Hope 2009; Ahanonu and Victor 2014)⁵.

Entertainment-Education Constructs

The extended elaboration likelihood theory (Slater and Rouner 2002), an extension of the elaboration likelihood model of persuasion (Cacioppo and Petty 1984), states that

⁵ Collective efficacy is particularly important for stigmatized behaviors perceived as generally accepted by one's community (Papa 2000). For example, many young and educated families continue to practice female genital mutilation as they fear their daughters will be eventually stigmatized in the marriage market (Hope 2009; Ahanonu and Victor 2014). A program that conveys the message that most community members are against this practice and would not retaliate against households who stop this practice could improve perceptions of controllability.

entertainment-education narratives can directly and indirectly influence behavioral health intentions. The actions of characters that audiences consider role models can directly help self-efficacious beliefs of audiences (Bandura 1977). For example, young audiences witnessing the death from AIDS of a favorite soap opera character may serve as a more powerful trigger for impacting intended and actual behavior change than a rationally-structured information-only message about the benefits of condom use (Airhihenbuwa 1999). While both types of messages may improve *behavioral beliefs* around the risks and consequences of unprotected sex in high-risk events, communications researchers argue that the emotionally-charged messages are more likely to motivate changes in intended and actual behaviors. This is particularly true when audiences *identify with characters*, as the message becomes more personal and relevant to their specific context (Cohen 2001; Singhal 2001). Character identification has several facets, such as perceived similarity, desire to be like the character, and believing to see things from the character's point of view (Murphy et al. 2011).

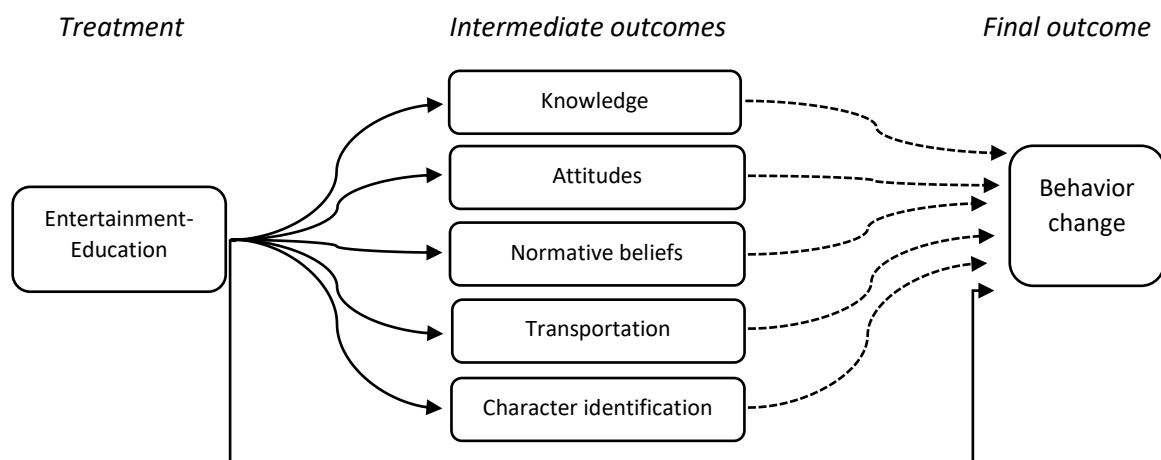
On the other hand, entertainment-education interventions may indirectly influence behavioral intentions through its impacts on psychosocial mediators. The extended elaboration likelihood theory (Cacioppo and Petty 1984; Slater and Rouner 2002) argues that when viewers are *transported or immersed in a program*, they are more likely to absorb embedded messages and change their beliefs and attitudes. Assuming impacts on psychosocial constructs of HIV knowledge, attitudes and normative beliefs translate to behavioral impacts, a program will be more effective in promoting the adoption of healthier behaviors the more it transports or immerses audiences in the program. Studies suggest that the key mechanism behind transportation is the reduction of counter-arguing, the thoughts that may dispute persuasive arguments (Benoit 1987). In other

words, audiences enjoying an entertainment program are less likely to question and rebut program messages. Unlike face-to-face HIV sensitization campaigns, audiences of entertainment media are less likely to perceive the educational or “preachy” objectives of the information source. In lab settings, Hall et al. (2011) and Moyer Guse (2008) found that increased transportation made audiences more receptive to positive messages by reducing counter-arguing.

If an entertainment-education intervention is effective in positively impacting the above psychosocial and entertainment-education mediators, these, in turn, should theoretically impact behaviors. Figure 1 depicts the theory of change of entertainment-education narratives theorized in this thesis. The moderators paper studies *if* the program worked for a population, and the mediators paper *how* it worked for a specific outcome. *While the former intends to measure the program’s impacts on a series of psychosocial and behavioral outcomes (solid lines in Figure 1), the latter measures the indirect effects that mediator gains had on increased HIV testing (dotted lines).* By decomposing treatment effects, the mediation analysis helps unpack the “black box” of how an intervention worked (Imai et al 2011). This process is often ignored by experimental studies, which have been criticized for paying little attention to causal mechanisms behind program impacts (Deaton 2010a and 2010b). Due to theoretical and pragmatical considerations, the mediation analysis was conducted only for the mediating variables of HIV knowledge, attitudes and transportation. A measure for normative beliefs was excluded because the program did not impact it, making this variable an unlikely mediator. Character identification was excluded from the mediation analysis for two reasons. First, this variable measures a psychological construct that is narrower than the other tested variables. Second, the scales of character identification were psychometrically

indistinguishable from the scales of the transportation variable (the correlation between both indexes was $r=0.76$), so the possibility that both mediators were measuring a similar construct could not be discarded. As a result, the effect size of both mediators on HIV testing was very similar.

Figure 1. Theory of change of entertainment-education narratives



The theory of change of entertainment-education narratives includes a series of implicit assumptions that need to be empirically tested. Knowledge gains will not materialize if audiences do not pick up the embedded messages or if they quickly forget them. A program is unlikely to improve a sense of self-efficacy if viewers consider the storyline or characters as unrealistic or its messages not applicable to their contexts. A program may have a limited effect on attitudes if it fails to change their perceptions of prevalent and accepted social norms.

Testing the Theoretical Framework from a Moderation Perspective

Assessing program moderators identifies whether the effect of an intervention varies for different groups⁶. From a moderation perspective, program effectiveness may be dependent on pre-intervention characteristics of the audiences. The Shuga trial hypothesized that exposure to an emotionally engaging entertainment-education drama would reduce risky behaviors for the average study participant. However, the direction and size of program impacts by gender and high-risk profiles are less predictable. For example, the portrayal of older men dating younger women could increase inter-generational sex among females, who are more likely to engage in this high-risk sexual behavior (Boer and Mashamba 2007; Balkus et al. 2015; Schaefer et al. 2017). Similarly, the portrayal of concurrent multiple sexual partners may potentially increase this behavior among males by rationalizing this behavior as “commonly practiced by others” (Hust, Brown and Engle 2008).

Randomized controlled trials (RCTs) can potentially provide unbiased estimates of program impacts on a series of outcomes. However, because the study was not designed, and the analysis was not pre-specified for these moderators, the experimental interpretation of the moderators paper should be read as hypothesis-generating warranting further confirmation in a subsequent study (Spieth et al. 2016).

Testing the Theoretical Framework from a Mediation Perspective

From a mediation perspective, the theory of change assumes that impacts on these potential mediators will promote safer sexual behaviors. However, experimental research

⁶ If the effects are the same, the intervention or treatment effect is called homogeneous. If the effects are different, they are called heterogeneous (Rotham and Greenland 2008).

in the field of behavioral economics has shown that the direction of these putative mediators is more complex than what rational behavioral models assume (World Bank 2015). For example, experimental research in Sub-Saharan Africa (Gong 2010; Baird et al. 2014) shows that HIV-positive people that were adversely surprised by the results of their HIV tests subsequently increased their risky sexual behaviors. When their impacts are in the expected direction, mediators may have a limited influence on the final outcome. A series of mediation studies of (non-entertainment-education) HIV behavior change interventions targeted at youth in South Africa (Jemmot et al. 2015; O’Leary et al. 2012) and the US (Baumler et al. 2012; Schmiede et al. 2009) suggest that while interventions had impacts on psychosocial variables (e.g., knowledge and attitudes towards condom use) and sexual behaviors, the latter were not mediated by the studied psychosocial variables. This highlights the importance of studying untraditional mediators that could explain the adoption of safer sexual practices in the context of RCTs. Although the main Shuga paper shows that the intervention was effective in impacting both the intermediate and final outcomes, it does not tell us which of these potential mediators influenced and influenced the most the adoption of safer sexual practices. Despite methodological improvements in identifying causal mediator estimates (Buis 2011; Daniel et al. 2011; Hicks and Tingly 2011), however, mediation analysis remains a correlational exercise for hypothesis-generating research (Gerber and Green 2012; Bullock, Green and Ha 2010). Complete mediation is expected to be rare in psychology because of the prevalence of multiple mediators (Preacher and Kelly 2011).

Current Gaps in the Literature

For the case of non-experimental field studies of HIV entertainment-education programs, important methodological limitations persist. Few quantitative studies properly address contemporaneous issues, as many field studies are based on before-after designs that lack comparison groups (e.g., Jadranin et al. 2015; Ramirez et al. 2014; Vernon et al. 2007). Selection bias (Heckman 1979) is likely to persist in studies based on matching techniques, as these frequently lack enough data to control for confounding factors when comparing program audiences with community members that decided not to watch the program (e.g., Boulay 2002; Collins et al. 2003; Kuhlmann et al. 2008). Similarly, experimental research is needed to study spillover effects on other community members to overcome selection issues of the current evidence (e.g., Geary et al. 2007; Mohammed 2001).

Experimental evidence of entertainment-education is nascent and comes from two strands of literature. The first comes from the communications literature. These studies are often efficacy trials aimed at understanding the pathways of entertainment-education mediators, such as character identification and transportation. However, these studies are greatly based on post-intervention surveys that are collected immediately after or within a few days of program exposure. Thus, in addition to potentially inducing self-reported biases due to observer or Hawthorne effects (McCarney et al. 2007), these studies are limited to measuring behavioral intentions and not actual behavior change. Cohort studies consistently find that the “intention-behavior gap” (i.e., the difference between intentions and actual behaviors) remains substantial, including for HIV testing (Mirkuzie et al. 2011), sexual partner concurrency (McGarrity et al. 2017) and general health behaviors (Sheeran and Webb 2016). Moreover, because this research is

conducted in lab settings and mainly conducted with US college students, its external validity⁷ (Campbell and Stanley 1963) is limited. Mediation studies are mostly based on one post-treatment round of data collection and observational data where theoretically it is not possible to rule out the reversal of the causal pathway (i.e., impacts on mediators subsequently affect the final outcomes, Baron and Kenny 1986; MacKinnon 1994).

The second strand of experimental research comes from the development economics literature. As part of the growing trend of RCTs in international development (Banerjee and Duflo 2009; Legovini et al. 2015), a series of field trials of entertainment-education interventions have been conducted across different development sectors. Field experiments of television and radio interventions have shown positive impacts on financial literacy behaviors among entrepreneurs in Nigeria (Coville et al. 2015) and the general population in South Africa (Gunhild and Zia 2013). Field experiments have shown that radio-listening groups were effective in increasing political participation among women in Liberia (Mvukiyehe 2017), and in reducing intergroup prejudice and conflict among listeners in Rwanda (Paluck and Green 2009). However, these interventions share a common feature: *the behaviors they aimed to change were publicly observed and therefore their results may not be easily transported or applied to (private) sexual behaviors.*

As shown in the meta-analysis study of this thesis (Orozco, Shen and Cluver 2018), a series of effectiveness trials have been conducted targeting youth sexual outcomes. Through

⁷ External validity, the extent to which the results of a study can be generalized to other settings or populations, may be improved by setting experiments in more “real-world” settings and using random samples from the wider population of interest (Cook and Campbell 1979). There are two types of generalizability: 1) *across situations*, the extent to which we can generalize from the study intervention to the “real-world” situation; and 2) *across people*, the extent to which we can generalize results from study participants to people in general.

two experimental trials of school-based interventions that included short video clips, researchers showed that teenage girls were less likely to engage in inter-generational sex in Kenya (Dupas 2011) and Cameroon (Dupas et al. 2017). Using an efficacy trial in Gabon, Milleliri, Krentel and Rey (2003) found that an educational comic distributed in schools was effective in increasing condom-use intentions. Through a repeated cross-sectional design where broadcasts of a popular radio show were randomized at the regional level, Vaughn et al. (2000) found that rural listeners adopted safer sexual practices in Tanzania. The meta-analysis concludes that effect sizes on reducing risky sexual behaviors were small, ranging between 0.07 and 0.15 standardized mean differences. Despite these small effects, its cost advantages are particularly important for reaching large segments of the general population (Wakefield, Loken and Hornik 2010), which is a policy imperative in countries with generalized HIV epidemics (UNAIDS 2007).

The validity of self-reported sexual behaviors has long been a limitation in the HIV literature (Padian et al. 2011), and the mentioned HIV field trials are no exception. These studies are based on self-reported outcomes and therefore are potentially exposed to social desirability bias (Brener et al. 2003; Minnis et al. 2009)⁸. Moreover, these HIV field studies of entertainment-education interventions do not address important knowledge gaps regarding program moderators and mediators. The target population (middle-school teenagers) or the sampling strategy (repeated cross-sectional versus panel data with pre-intervention individual-level data) prevent them from studying program effectiveness on key subpopulations that may be driving the HIV epidemic, such as individuals that

⁸ Social desirability bias is a type of response bias that takes place when survey respondents answer questions with the objective to be viewed favorably by the survey enumerator. In our study, respondents may under-report high-risk sexual behaviors as they may perceive them as undesirable behaviors (and over-report “good” behaviors, such as HIV testing).

regularly engage in high-risk sexual behaviors. On the other hand, there remains little research examining the influence of psychosocial and entertainment-education mediators in field settings. While the number of field experiments of this theory-based approach has grown in the last decade (DellaVigna and La Ferrara 2015), mediation studies remain rare outside the psychology literature due to confounding concerns of mediation estimates (Bullock, Green and Ha 2010; Gerber and Green 2012). None of the mentioned field experiments that targeted sexual or non-sexual outcomes conducted a mediation analysis.

Thesis Contributions to the Literature

This thesis addresses three important knowledge gaps of the HIV entertainment-education literature.

Meta-analysis of entertainment-education (Orozco, Shen and Cluver 2019)

Previous systematic reviews of mass media interventions that targeted sexual behaviors (Lacroix et al. 2014, Shen and Han 2014) included studies with a series of methodological constraints (e.g., selection bias, short-term surveys) that limit their ability to provide reliable impact estimates. The heterogeneity of their studies was also substantial. For example, previous reviews included lab studies; information-only media interventions that lacked a narrative format (e.g., flyers, public service announcements); or interventions with family planning messages that targeted married couples. By restricting the meta-analysis to narratives and to field interventions with HIV-messaging targeted to unmarried youth (i.e., homogenizing the intervention and population), Paper 1 aims to

better identify the effectiveness of the storytelling format on HIV outcomes for youth in a field setting. To the best of the candidate's knowledge, this is the first review of entertainment-education that includes high-quality studies that meet the Cochrane Review standards and the first to focus on interventions that followed the narrative format.

Impacts on HIV knowledge, attitudes and behaviors for high-risk youth (Orozco 2019)

Studying impacts among high-risk individuals per gender addresses important policy questions. Case et al. (2012) estimated that sexual concurrency represented over a third of new HIV infections in Sub-Saharan Africa. Population surveys from this region show that compared to women, men are less likely to get tested for HIV and therefore to start treatment (Cohen et al. 2011; UNAIDS 2016). These adverse behaviors combined with a larger sexual network increases the potential of high-risk males to infect more sexual partners (Banerjee et al. 2015; Padian et al. 2011; Jones et al. 2014). Similarly, promoting safer sexual practices among high-risk females can have population-level effects in Sub-Saharan Africa. Population surveys show that young women aged 15-24 consistently have higher HIV prevalence rates and acquire the virus 5-7 years earlier than their male peers (Joint United Nations Programme on HIV/AIDS 2014).

Informed by considerations relevant to the HIV epidemic in Sub-Saharan Africa, this study defined (in a post-hoc manner) high-risk individuals as those individuals that at baseline reported in the last six months at least one of the following four HIV-risk factors: (i) *Had unprotected sex in their last sexual encounter*. Unprotected sex remains the main source of HIV transmission among youth in Nigeria and Sub-Saharan Africa (Kharsany et al. 2016;

UNAIDS 2016; UNICEF 2017); (ii) *Engaged in sexual concurrency* (i.e., simultaneously had two or more sexual partners). This high-risk behavior increases the likelihood of contracting the virus per sexual event as recently-infected individuals are substantially more contagious than those who have had the virus for over six-eight weeks (Epstein and Morris 2011, Koopman, Simon and Riolo 2005); (iii) *Engaged in inter-generational sex*, defined as having had sexual partners ten years younger or older (Balkus et al. 2015; Schaefer et al. 2017)⁹. Research suggests that inter-generational sex between younger women and older men is an important driver of the HIV epidemic in Sub-Saharan Africa as older men tend to have substantially higher HIV prevalence rates in a context of low condom use (Poulin 2007; Dupas et al. 2011; Toska et al. 2015; Schaefer 2017); and (iv) *Reported having symptoms for a sexually transmitted infection (STI)*. High-risk sexual behaviors that lead to STIs also lead to HIV infection. Medical models suggest that active genital sores from other STIs can plausibly facilitate the transmission of HIV, though research has failed to confirm this hypothesis (Ward and Ronn 2010; Ng et al. 2011).

As secondary analysis, the moderators paper also explores the effect-modification of religion and television-consumption levels among high-risk individuals. The research of both potential moderators remains restricted to lab settings or to observational data in field settings (Sherry 2002). It was theorized that impacts would be lower for Muslim viewers as Shuga's messages of safe sex rather than abstinence adhere more to Western values (Peterson and Donnenwerth 1998; Collins and Sturgill 2013). Similarly, it was hypothesized that effects would be lower for individuals that watched television every

⁹ The HIV literature commonly classifies age-disparate relationships with ≥ 10 years age difference as inter-generational; for 5–9 years differences as intra-generational; and for 0–4 years differences age-homogeneous.

day, due to theoretical saturation effects of media consumption (Heather et al. 2008). To the best of the candidate's knowledge, the moderators analysis is the first study of an HIV entertainment-education narrative to study program effectiveness for high-risk individuals based on a RCT; and the first to use objective measures of HIV testing, condom-demand and STI biomarkers for an entertainment-education intervention.

Mediation analysis for HIV testing (Orozco and Cluver 2019)

The mediation analysis provides indicative evidence of three high-level theoretical mediators on increased HIV testing. This behavioral outcome has received renewed attention given experimental evidence that shows that starting HIV treatment early in the disease is an effective way to reduce new HIV infections (Cohen et al. 2011). Despite important investments to increase access to HIV testing in Nigeria (National Agency for the Control of AIDS 2015), it is estimated that only 34 percent of HIV-positive people are aware of their HIV-status (UNAIDS Data Book 2017)¹⁰. Only one in eight sexually-active young adults have been tested in the last 12 months and knew their status (UNICEF report, Multiple Indicators Cluster Survey 2016-2017).

The mediation research addresses an important knowledge gap with respect to *motivating* safer sexual practices, considering that (i) HIV awareness is already high in the region and in Nigeria¹¹; and (ii) evidence of the limited effectiveness of HIV-information campaigns in Sub-Saharan Africa (Harrison et al. 2010, Michielsen et al. 2010, Mavedzenge

¹⁰ This percentage is considerably below the 90-90-90 targets of the UNAIDS Fast Track Strategy (UNAIDS 2014). By 2020, this target intends for 90 percent of people living with HIV to get tested and learn their status; of whom 90 percent are to be on treatment; of whom 90 percent are to be virally suppressed.

¹¹ According to 2013 Nigeria Demographic Health Survey, among urban 15-24-year-old Nigerian youth, 98 percent had heard about HIV/AIDS and 80 percent knew that condom use prevented HIV infection.

et al. 2011, Krishnaratne et al. 2016). *The mediation analysis intended to understand the absolute (in percentage points) and relative (as percent of the total treatment effect) influence on increased HIV testing of the mediator variables of HIV knowledge, attitudes and transportation.*

The mediation analysis selected HIV testing from a list of behavioral outcomes for three reasons. First, encouraging youth to get tested for HIV is one of the two main messages of MTV Shuga (reducing risky sexual behaviors is the other). Second, items related to HIV testing were included in the measures for both psychosocial outcomes. Third, impacts on HIV testing were statistically significant for two different periods, allowing researchers to measure indirect or mediating effects in the short- and medium-term. Moreover, the medium-term indicator was independently observed by researchers, overcoming concerns of social desirability bias of self-reported measures (Brener et al. 2003; Minnis et al. 2009).

To the best of the candidate's knowledge, the mediators paper is the first field study of an HIV entertainment-education that measures potential mediators (both psychosocial and entertainment-education variables) beyond two weeks from program exposure; and the first to use an objective measure for measuring the behavioral outcome. This paper is also among the first to study time-decay relationships of mediators.

Understanding different program effects of the Shuga trial

Figure 2a shows the logic model (Alter and Murty 1997) of the Shuga trial. It is theorized that individuals that watch MTV Shuga in community screenings (*inputs*) will learn a

subset of the embedded messages that will improve their HIV knowledge and attitudes, which, in turn, will subsequently increase the adoption of safer sexual practices. It is hypothesized that the program's impacts should be higher for individuals that were transported or immersed in the program.

While impacts or gains in knowledge, attitudes and behavioral intentions should occur immediately after program exposure (*outputs*), the logic model assumes that impacts on knowledge, attitudes and realized behaviors should persist 8 to 10 months after program exposure (*outcomes*), when the post-treatment data collections take place. The increase of safer sexual practices should potentially be reflected in the reduction of HIV infections and other STIs among youth (*impacts*).

Figure 2a. Logic Model of the Shuga trial

Inputs	Outputs	Outcomes	Impacts
Community screening of entertainment-education program. MTV Shuga is an entertainment-education intervention aimed at improving knowledge, attitudes and behaviors regarding HIV/AIDS of African youth audiences. The Nigerian NGO "Breaking Barriers In Development" carried out 240 community screenings in both treatment and control locations. NGO staff was explicitly asked not to provide additional information, especially regarding HIV/AIDS resources. Most of the screenings took place in schools (these were closed as the screenings took place over the weekends), while others took place in community centers. Study participants were invited to three 50-minute	Study participants liking and being emotionally engaged by MTV Shuga. Immediate impacts on attitudes and behavioral intentions	Improvements in the psychosocial outcomes of (1) knowledge about HIV transmission and benefits of HIV testing and (2) more favorable attitudes towards HIV-positive people. Behavioral responses: - Increase in HIV testing - Decrease of sexual partner concurrency, unprotected sex with main and other sexual partners, and inter-generational sex - Decrease of sexually transmitted infections	Decrease in HIV infections and other STIs among Nigerian youth

entertainment community screenings, for a total of 2.5 hours. The three sessions were conducted in three consecutive weekends. The full third season was shown without being changed. The NGO greatly adhered to the planned treatment intervention. A field-based coordinator supervised the implementation of screenings.	<i>(outputs were measured immediately after the community screenings)</i>	(STI) due to safer sexual practices <i>(all outcomes were measured 8 to 10 months after the screenings)</i>	
---	---	---	--

Figure 2b shows the conceptual framework for studying variation in program effects, treatment contrasts and implementation as suggested by Weiss, Brock and Bloom (2013). To replicate as much as possible the television-broadcasting setting, a placebo rather than a pure control was used to “net-out” the effects of the program’s delivery mechanism (i.e., the community screenings). The upper row (in *Italics*) represents the treatment condition of this randomized control trial (i.e., exposed MTV Shuga) and the lower row represents the control condition (i.e., exposed to the placebo television program Gidi Up).

The quality of *treatment fidelity*, the difference between the planned and offered treatment (the difference between items 1 and 2), was high. Supervised by research coordinators in the field, the implementing NGO greatly followed the implementation guidelines developed by the research team. Invitations to and the delivery of community screenings were successfully implemented in all 80 study locations with no major implementation issues reported. As planned, both experimental groups were treated in a similar way with the only difference being what each community center screened. The quality of *treatment contrast*, the difference between what the treatment and control groups received (the vertical difference of items 3), is high given the full adherence of the community screenings. While locations that were assigned to the treatment condition only screened MTV Shuga, locations that were assigned to the control condition only screened the placebo program. *Program uptake*, the difference between treatment

offered and treatment received at the study participant level (the difference between items 2 and 3) was high, with over 85 percent of invited individuals attending both the treatment and the control community screenings. Finally, given the high-quality implementation of the studied interventions and the low attrition rates of the panel data (under 3 percent), the difference between the mean outcomes of the treatment and the control groups (vertical difference between items 5) provide a high-quality counterfactual for studying program effectiveness of MTV Shuga.

Figure 2b. Conceptual framework for studying variation in program effects, treatment contrasts and implementation

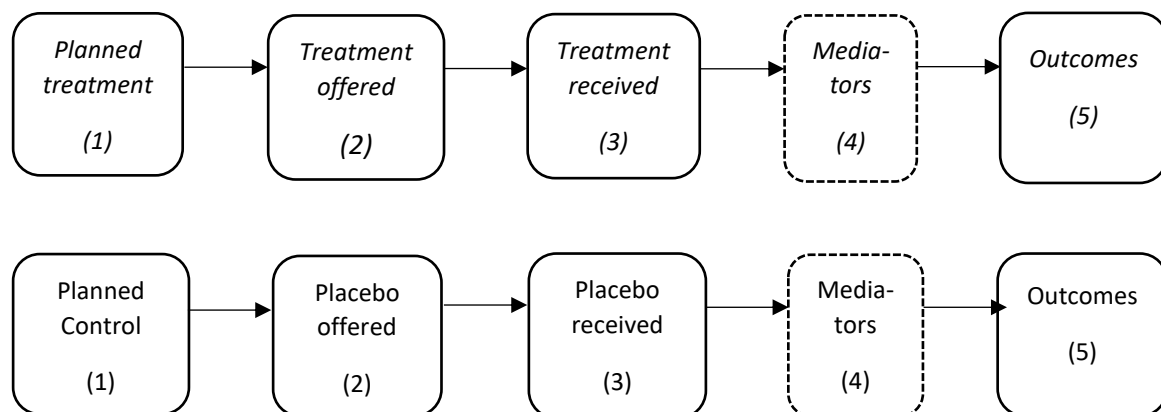


Figure 2b also highlights an important strength of the mediation analysis, to be expanded in the Methods chapter. In addition to meeting the time-sequence requirement of mediation frameworks (i.e., impacts on mediating variables take place after program exposure and prior to the behavior, observed in Figure 2b with item 4 preceding item 5), the experimental design of the Shuga trial provides a high-quality control group. The latter removes one of two potential bias sources when estimating mediation quantities (i.e., confounding factors for program estimates on intermediate and final outcomes).

CHAPTER 2. METHODOLOGICAL FRAMEWORK

This chapter presents the methodological frameworks upon which subsequent chapters of the thesis are based. The first part presents the methodological framework of the meta-analysis, discussing its inclusion criteria, search strategy, risk of bias and quality of evidence grading, and data analysis. The second part describes the methodological framework of the Shuga trial, including the trial's setting, recruitment strategy, ethical considerations, and evaluation design. This section then discusses the measures and the statistical analysis used in the moderators and mediators papers.

The moderation analysis follows the 2010 Consolidated Standards of Reporting Trials (CONSORT) statement (no equivalent guidelines exist for mediation analyses). For example, the thesis provides detailed information regarding the trial's eligibility criteria and recruitment strategy, of its treatment and control interventions, and the randomization and statistical methods used. The thesis includes the CONSORT flow diagram, tables with baseline demographic and clinical characteristics for each group and discusses the trial's limitations and the generalizability of results. The moderators paper complements the CONSORT guidelines with the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) and the Guidelines for the Content of Statistical Analysis Plans in Clinical Trials (Gamble et al. 2017). For example, this paper notes in its title and text that the analysis is post-hoc (i.e., not pre-specified), provides details to the trial's registry, the appendices provide extensive details regarding the research ethics approval and consent process, the time points at which the outcomes were measured, provides details about all measures used in the analysis, and describes the roles, responsibilities and affiliations of all co-authors.

Meta-analysis of entertainment-education narratives (Orozco, Shen and Cluver 2018).

Inclusion Criteria

Following Cochrane Review standards, studies were eligible for inclusion if they were RCTs, cluster RCTs, and quasi-experimental studies with robust counterfactual designs, including interrupted time series designs with control groups. To focus on changes in realized and not intended behavior change, the meta-analysis included studies that measured outcomes at least one week after exposure to an intervention. Studies of complex interventions (those that included other information components) were included if the edutainment component played a prominent part. This review follows reporting standards for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al 2009). This study was registered in the PROSPERO International prospective register of systematic reviews (PROSPERO 2016: CRD42016046005). The registration form of the systematic review can be found in Appendix 1.

Search Strategy

Published studies were searched in English. The search period covered 1985 to the latest available year. Searches were completed on March 31, 2017. The following databases were searched: Communication and Mass Media Complete, EBSCO, ERIC (CSA), EconLit, JSTOR, MEDLINE, CINAHL (EBSCO), JSTOR, metaRegister of controlled trials, NBER, OpenSIGLE (grey literature), PsycINFO, RePEC, and the Cochrane Library. References from previous reviews and from retrieved articles were examined for additional studies. Journal articles published reports and, for the case of economics research, published working

papers were included (prior to journal publication, the latter are standard practice in the economics literature). Authors were contacted for further information when required.

Risk of Bias assessment and Quality of Evidence

Study quality was assessed using the Cochrane Review guidelines for assessing risk of bias (Higgins and Greene 2011). Each study was evaluated for participant selection (sequence generation and allocation concealment), blinding (personnel and outcome assessors), attrition (> 80% retention), and selective reporting and other types of biases. Each item was scored as either low, high or unclear risk of bias.

The quality of the evidence was graded using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria (GRADE Working Group 2004).

Data Analysis

For cluster RCTs or natural experiments delivered that failed to cluster their standard errors; the meta-analysis used a conservative value for the intra-class correlation coefficients (ICC) of 0.05 (Glassman et al. 2015)¹². The analysis preferred intent-to-treat rather than per-protocol estimates to minimize selection issues. Standard meta-analytical methods were used. The analysis used the Comprehensive Meta-Analysis software to convert all binary and continuous data outcomes to standardized mean differences and their 95% confidence intervals (DerSimonian and Laird 1986; Wilson and Lipsey 2001). When required, signs of study estimates were changed with positive values to represent

¹² The similarity of individuals within clusters reduces the variability of responses in a clustered sample, eroding the power to detect true differences between study arms. The intra-cluster correlation coefficient, which expresses this similarity, is used with the cluster size and the number of clusters to calculate the effective sample size in a clustered design.

safer sexual practices. For studies that reported different types of primary behaviors, outcome values were aggregated to four behavior categories: (i) age-gap with sexual partners; (ii) number of sexual partners (iii) unprotected sex; and (iv) STI testing and follow-up.

The primary analysis for each of the four behaviors includes some degree of heterogeneity as it synthesizes data from non-identical interventions and from similar but different outcome indicators. Study heterogeneity is recognized by applying the random-effects model (Hedges and Vevea 1998) when pooling the results and by calculating the I^2 statistics (Higgins and Greene 2011; Higgins et al. 2003). The random effects model estimates the mean of a distribution of true effects and because each study is estimating a different effect size, the weights assigned to each study are more balanced than in the fixed effects model. The I^2 statistic describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error. Because factors co-vary across subgroups in meta-analyses and specific limitations for this study described below, the subgroup analysis is described as hypotheses-generating rather than hypotheses-testing (Spieth et al. 2016).

To determine whether results were robust to methodological decisions, the meta-analysis re-ran the analysis for the primary (behavior) outcomes by repeating the meta-analysis using fixed-effects models; removed one study at a time to determine if it substantially affected the meta-analyzed effect size; and re-ran the analysis without studies that were rated with high risk-of bias. This study assessed whether publication bias was present through funnel plots (Egger et al. 1997) and Rosenthal's (Rosenthal 1979) and Egger's tests (Sterne et al. 2001). While the former test is a graphical test, the latter provides an

estimate for the number of new studies that would need to be included for the effect size to be statistically different.

The meta-analysis is the first paper of the thesis and it can be found in Chapter Three.

Moderators and Mediators Analyses of the MTV Shuga Trial

Research Setting

Impact evaluations estimate the impact that can be causally (and fully) attributed to a development program or intervention. Their main challenge is identifying a control group that is comparable to the treatment group. RCTs are typically considered the “gold standard” design, since by construction treatment assignment is uncorrelated with other characteristics of the treated population (i.e., the estimated effect is not confounded). When take up and attrition rates are not systematically different between experimental groups, causal estimates of program impacts should be unbiased (Glennerster and Takavarasha 2013).

In 2013, the doctoral candidate launched the Shuga trial with an initial grant from the Bill and Melinda Gates Foundation. In 2014, the candidate secured additional funding to expand the Shuga trial, subsequently hiring and supervising the broader research team, the survey firm (Hanovia Medical Limited), the Department of Medical Microbiology and Parasitology (University College Hospital, Ibadan) to collect biomarkers for STIs and other objective measures, and the implementing NGO that carried out the 240 community screenings (Breaking Barriers in Development). The candidate had worked with most of

these organizations in previous RCTs in Nigeria¹³. The Shuga trial was registered at the American Economic Association RCT Registry (<https://www.socialscisceregistry.org/trials/509>).

The Shuga trial took place in Ibadan, the third largest city in Nigeria, and six surrounding towns in the southwest region. The study sites were in the states of Oyo, Osun and Ondo. Figure 3 displays the geographical location of these three states. Choosing states that were contiguous facilitated the logistical implementation of the trial and reduced transportation costs. The epidemic in the study location is generalized, defined by UNAIDS (2005) as exceeding 1 percent of the general population. The southwest region has an adult HIV prevalence rate of 2.8 percent, just below the national average of 3.1 percent (Nigeria Global AIDS Response Report 2017). This region was selected based on four criteria. First, most urban youth speak English, the language of MTV Shuga¹⁴. Second, the region was not affected by security issues that could have affected the integrity of the study. Third, the southwest region is demographically diverse, with a relative balance between the main religious groups of Nigeria: Christians and Muslims. The study pilots suggested that “female seclusion¹⁵” was not an issue in the study sites, which minimized concerns of take-up rates differing for men and women. Fourth, this region has relatively weak television-broadcasting signals, minimizing concerns of Shuga broadcasts (the third

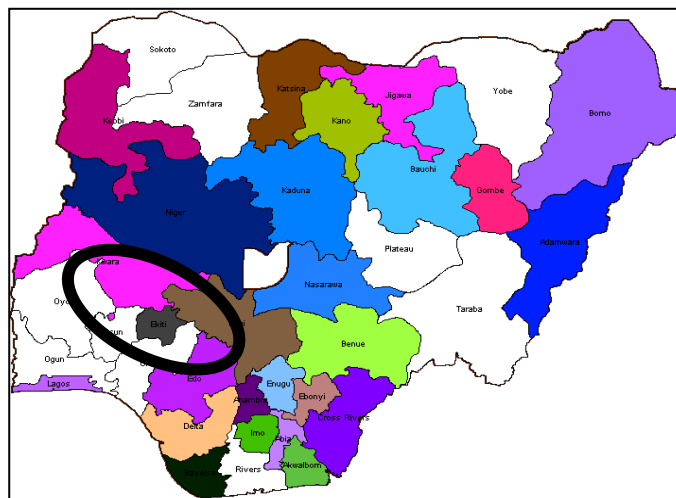
¹³ In early 2018, the candidate launched, with these same organizations and Discovery Learning Alliance (the foundation of the Discovery Channel), a new trial in northern Nigeria. The trial will study a social norms campaign combined with mobile-based interventions aimed at increasing educational outcomes.

¹⁴ A region with many national universities, this region was a priority for large marketing campaigns of MTV Staying Alive Foundation (e.g., college concerts, media advertisements). The doctoral candidate agreed with this organization postponing their marketing campaigns for the duration of the Shuga trial.

¹⁵ Female seclusion refers to various practices in traditional Muslim societies that confine women to the company of other women or close male relatives in their homes or in separate female living quarters. Female seclusion greatly restricts women participating in social activities outside the household (The Oxford Dictionary of Islam 2019).

or previous seasons) watering-down estimates of program effects. Although television broadcasts would have unlikely biased impact estimates due to the randomization strategy, television broadcasts could have potentially lowered the effect size and, therefore, the statistical power of the Shuga trial.

Figure 3: Study states in southwest Nigeria



Study Sample

The sample was designed to be representative of the general urban youth population, in view of the mass media nature of the intervention. The trial invitations to potential study participants targeted by 18-25-year-olds and were stratified per gender (50 percent female) and oversampled individuals in relationships (two-thirds). Other than age, no further exclusion criteria were used at the invitation stage.

Underage youth were not included in the trial for two reasons. First, their inclusion would have reduced the study's statistical power. The median age of sexual debut in southwest Nigeria was 19.8 years of age (2008 NDHS). Second, the inclusion of underage youth would have required parental consent at the invitation stage, which in turn could have adversely

affected study findings and their generalizability. Parental consent would have required explicitly disclosing the educational objectives and sexual-content of MTV Shuga to both parents and study participants. This could have potentially triggered Hawthorne effects (McCarney et al. 2007) among the treatment group (as the placebo lacked sexual messages), affecting the study's internal validity (Levine and Parkinson 1994)¹⁶. In addition, parental consent would have likely reduced the number of youths from conservative households participating in this study, affecting the study's external validity (generalizability across people, Cook and Campbell 1979).

The listing data (the sampling framework collected prior to the baseline survey) collected information on gender, age, education and relationship status. Analysis of the listing data suggests that study participants were similar in observable characteristics to individuals who decided not to attend study screenings, indicating that study participants were greatly representative of invited youth. The first column of Table 1 presents the pre-intervention characteristics of the Shuga sample. The sample's average age was 20.6 years and the sample of study participants remained balanced by gender (47 percent female). Two thirds were Christian and one third Muslim. Baseline respondents had an average of 11.6 years of education, twice the national average.

The baseline data shows important challenges related to attitudes towards HIV-positive people, and high-risk sexual behaviors among youth. Six in ten respondents said they would not buy fresh vegetables from an HIV-positive shopkeeper. Half reported not using

¹⁶ Internal validity, the extent to which a causal finding is warranted, is determined by the degree to which a study minimizes systematic error or bias between the dependent and the independent variable. By randomizing treatment assignment (independent variable), experimental studies are better positioned to attribute the observed differences in the outcome (dependent) variable to the treatment than non-experimental studies.

a condom in their last sexual encounter, despite the fact that less than nine percent of respondents were married. Only a quarter of baseline respondents reported having ever gotten tested for HIV. The second column of Table 1 presents the population-level estimates for the southwest region derived from the microdata of the 2008 Nigeria Demographic and Health Survey. Despite important sampling differences (e.g., the NDHS survey had many more females and included rural areas), the Shuga sample was similar to the region's averages in terms of religion (a third Muslim and two thirds Christian), years of education (around 11 years) and television-ownership rates (nine in ten households).

The modules of HIV knowledge, attitudes and behaviors of the Shuga instrument included items of the 2008 NDHS instrument, allowing the comparison of both samples. With respect to the population-level sample, the attitudes of the Shuga sample shows no clear trend in terms of being more (or less) accepting of HIV-positive people: while study participants would be more likely to keep the HIV status of a family member a secret and less willing to buy from an HIV-positive shopkeeper, study participants were also substantially less likely to blame an HIV-positive person for his disease. Similarly, no clear pattern of higher (or lower) levels risky sexual behaviors is observed with respect to the region's averages. Whereas study participants were less likely to report ever testing for HIV and to report a larger number of sexual partners, they were also more likely to report using a condom in their last sexual encounter. The lack of exclusion criteria for the Shuga sample (other than age), combined with the observed balance between study participants and invited youth, and its similarity with respect to the region's average characteristics and HIV outcomes, suggests the Shuga sample is to some extent representative of the urban youth population in the study region.

Table 1. Comparison of Shuga sample with urban population-level survey of SW Nigeria

Variable	Shuga Baseline		DHS - South West	
	N	Mean	N	Mean
Socio-demographics				
Age	5,166	20.60	2,186	21.39
Gender (Female =1)	5,166	0.47	2,186	0.68
Christian	5,166	0.63	2,186	0.67
Muslim	5,166	0.37	2,186	0.33
Years of education	5,166	11.60	2,186	10.62
Household owns a television	5,166	0.91	1,475	0.87
House has cement walls	5,166	0.97	1,475	0.82
HIV Attitudes				
If a member of your family got HIV, would you want it to remain a secret?	5,166	0.68	1,961	0.55
If you knew that a shopkeeper had HIV/AIDS, would you buy fresh vegetables from him/her?	5,166	0.42	1,961	0.55
People with HIV should be blamed for bringing the disease into the community.	5,166	0.36	1,961	0.64
Sexual behaviors				
Used condom last time	2,678	0.51	1,396	0.36
Ever tested for HIV	5,155	0.26	1,961	0.37
Number of sexual partners in the last 6/12* months (among those who had started having sex)	2,933	1.24	1,396	1.11

N= Sample size, Mean=Average or proportion for continuous and binary variables respectively. *The Nigeria DHS used a recall period of 12 months, while the Shuga trial used a recall period of 6 months. The latter was determined by the timing of the follow-up survey, originally planned six months after the intervention was conducted.

THE STUDY INTERVENTION

MTV Shuga is a theory-based educational drama broadcasted in over 70 countries, including all Sub-Saharan countries (MTV Staying Alive Foundation website, September 2015). Featuring prominent African actors and singers, the television series seeks to address the spread of HIV/AIDS by fusing sexual health messaging with engaging

characters and storylines¹⁷. The trial studied the third season, which took place in Lagos with most characters in their mid-twenties.

The third season lasts less than three hours (composed by 8 episodes of 22 minutes each), making it a relatively “low-dose” intervention considering that soap operas often last months or years on the air. No further HIV information was provided to study participants to focus on the effectiveness of the television drama alone. The program was produced by MTV’s Staying Alive Foundation and financially supported by African governments, UNICEF, PEPFAR, the Elton John AIDS Foundation, and The Bill and Melinda Gates Foundation, among others. MTV Shuga primarily aims to increase HIV testing and reduce risky sex (versus promoting abstinence). See Appendix 3 for scripts and main HIV/AIDS messages.

To achieve its behavioral objectives, this theory-based intervention embeds educational messages that aim to first impact key psychosocial variables. For example, specific storylines intended to *educate audiences about HIV transmission and HIV testing* (e.g., the need to get tested again after three months of potential HIV exposure); to favorably change their perceptions about *prevalent social norms* (e.g., youth should know better and not stigmatize HIV-positive people like older people often do); and through empathy with characters, to positively change audiences’ *attitudes towards HIV-positive people* (e.g., Weki, a teenage boy born with HIV, is stigmatized by the parents of his friends when he tries to remain in a soccer league).

¹⁷ According to one of its producers, by changing the character and plot when delivering a given message, the latter can be repeated multiple times without audiences perceiving the message as too repetitive or too “preachy”. This, in turn, increases the time of exposure to program messages, which is an advantage over explicitly educational interventions, such as school sex education and HIV community sensitization campaigns.



Weki's teammates find his anti-retroviral (ARV) treatment in his locker, and he is forced to reveal that he HIV-positive. His teammates don't say anything then, but before a big match, their parents withdraw all of them from the pitch, asking the coach to kick Weki out of the team. Weki's teammates, acting against their parents' will, decide to play the match. However, Weki, at that point, realizes that his HIV status is making life complicated and

decides not to continue playing football.

The Shuga trial involved community film screenings in 80 urban and semi-urban locations. The 80 locations were defined as two-mile circles centered around a community center (often a school) where the screenings took place over the weekend. The trial invited individuals living near these centers to attend entertainment film screenings (i.e., the invitation and the intervention did not include further HIV information).

Three screenings were hosted in each location (for a total of 240 community screenings) over a period of four weeks. The first screening (referred to as screening "zero") served to recruit potential study participants, while the latter two screened the treatment and control programs. Screening zero was the same in all locations, as the intent was to recruit amongst those interested in attending community screenings in general. This induced-selection also helped increase take up rates for subsequent screenings. In treatment and control locations, attendees of screening zero were invited to attend two more film screenings. To increase take up for attending all three screenings, study participants were provided with lottery tickets for attending that had mobile airtime as the price (see invitation ticket in the below picture). In screenings two and three, Shuga was screened in treatment locations and a placebo drama in control locations. The invitation and the

implementation of the treatment and control community screenings were delivered in the same way, including in their screening time (just under three hours).

With the advice of Nigerian colleagues, the research team selected the television series “Gidi Up” as the placebo program. This is a commercial high-quality production that follows a similar storyline to MTV Shuga (i.e., Gidi Up follows the life of four young creatives who are pursuing success and independence in Lagos), with 25-minute episodes (versus 22 minutes). Unlike Shuga, the placebo series lacked HIV educational messages. To avoid potentially inflating results through the placebo having negative effects on sexual outcomes, the control episodes were selected to minimize storylines heavy with sexual plots. See Appendix 4 for the Template for Intervention Description and Replication Checklist (TIDieR), which provides further information about the design and the implementation of the treatment and control arms of the Shuga trial.

Picture: Invitation ticket and community screening



Other evaluation alternatives considered

The doctoral candidate considered two alternatives before selecting community screenings. Both alternatives considered using television broadcasts, closer to the “natural” delivery mechanism of the main media outlet of this drama. However, it should be noted that Nigerian urban youth are increasingly watching “television” programs through other media outlets. According to a marketing study (Consumer Lab Ericsson 2015), 57 percent of 18-24-year-old youth wanted to see television programs at their leisure, with two-thirds of the viewing hours conducted through laptops, mobile phones and tablets.

The first option was a pipeline difference-in-difference strategy where media markets would be randomized into two broadcast waves (i.e., the first wave would serve as the treatment and the second wave as the control). A similar strategy was used in natural experiments¹⁸ by La Ferrara et al. (2013) and Kearney and Levine (2014) to study sexual outcomes for Brazilian adults and American teenagers respectively. However, this difference-in-difference strategy was considered unfeasible after the candidate conducted field meetings with the Nigerian telecommunications agency in 2013. Government officials noted that they could not guarantee that local broadcasters would respect the randomization plan, raising concerns of group contamination and violations of internal validity (Levine and Parkinson 1994). Additionally, conversations with Nigerian research colleagues suggested that the first two seasons of MTV Shuga were not very popular in Nigeria. Low viewership rates would have led to low take-up rates of the

¹⁸ A natural experiment occurs when individuals or clusters of individuals are exposed to the experimental conditions resembling random assignment. Exposure is determined by other factors outside the control of researchers.

television broadcasts and therefore a weak treatment instrument (Gerber and Green 2012). Low viewership rates were confirmed by the baseline survey: 2 percent of study participants had watched previous seasons of MTV Shuga.

The second alternative considered was a symmetric encouragement design. This identification strategy was used by Gundhild and Zia (2013) in the context of embedding financial literacy messages in *Scandal!*, a popular television drama in South Africa. While treated individuals would be financially encouraged to watch *Scandal!*, control individuals would be financially encouraged to watch another television program broadcasted at the same time in a different channel. Payments would be conditioned on correctly responding to content questions by automatic telephone calls. Although this approach seemed to be operationally feasible in major cities of Nigeria, it had two important shortcomings. First, in view of the very low viewership rates of the show, financial incentives would need to have been substantial to achieve the required sample size. Second, (such large) conditional payments would have likely generated important Hawthorne effects (McCarney et al. 2007) among study participants (i.e., people paying additional attention to the television programs than they would normally do knowing that payments were conditional on their correct responses), potentially lowering the internal and external validity of the study. The candidate is carrying out a symmetric encouragement design to study a mass media campaign to promote solar lanterns in Senegal (Coville, Orozco and Reichert, 2019).

For the above reasons, the doctoral candidate decided to use community screenings as the delivery mechanism for studying program effectiveness of MTV Shuga. Moreover, greater logistical control over the community screenings allowed experimentally testing a

series of additional treatment arms. However, the community-screening setting lowered the external validity of the study in two ways: departing from the “natural” television broadcasts of the intervention at scale (generalizability across situations) and by pre-selecting people who were willing to attend film screenings (generalizability across people). Moreover, the study design probably overestimates the effectiveness of the show compared to when watched on television or online. For example, audiences at home interested in these types of shows may be more likely to get distracted or less likely to watch the full season.

EVALUATION DESIGN

The Shuga trial was a cluster RCT that also contained an individual-level randomization component. Table 2 summarizes the evaluation design. The 80 locations were randomly assigned to two treatment arms and one control group. The Shuga trial studies two channels through which the drama may have worked. The *individual channel* measures the effects of program exposure. The *social channel* is the program’s effects on viewers’ perceptions of social norms, using local peers or friends as local references groups (the show, however, also conveys social norm information for the “broader” group of national and regional audiences).

Treatment 1 (T1) captures the individual channel. The trial intended to experimentally study the social channel, or how conforming to social norms may subsequently affect the sexual behaviors of study participants (i.e., an experimental mediation analysis), through two notions of reference groups.

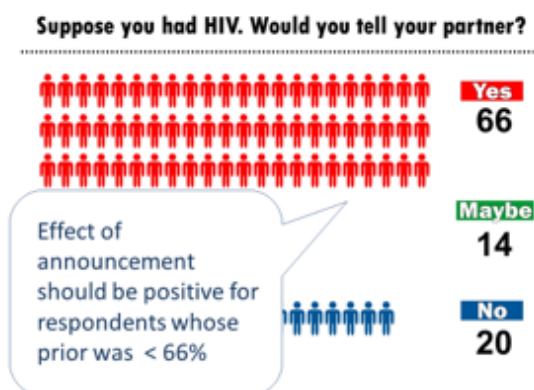
Table 2. Evaluation design of the MTV Shuga trial

	T1 (1/3 locations)	T2 (1/3 locations)	Control (1/3 locations)
Attends without Friends	Shuga screening (1/6 participants)	Shuga screening + Video clips & Graphs (1/6 participants)	Placebo screening (1/3 participants)
Attends with Friends (T3)	1/3 individuals randomly selected to receive 2 extra tickets (1/6 participants)	1/3 individuals randomly selected to receive 2 extra tickets (1/6 participants)	

The first reference group were local peers of similar age and background, a generic reference group. Their potential influence was measured in the following way. In half of the treatment locations (T2), the community screenings added a 5-minute video-clip containing interviews of local peers after watching Shuga and “smart graphs” with statistics about their social norms. The recorded interviews and the collection of this data was conducted during the piloting of the community screenings (in locations near but outside the 80 study locations).

Box 1 presents an example of one of these graphs. It was hypothesized that participants’ beliefs and attitudes would converge towards these social reference-points.

Box 1. Example of Smart Graph used for Treatment 2 of broader Shuga trial



The research team expected the effects of the announcement to be positive for respondents whose prior beliefs were less than 66% percent. Moreover, we expected stronger effects for individuals whose initial priors were furthest from the announced reference points. In other words, if an individual's prior beliefs came close to the provided reference points, his priors would not be updated (i.e., he does not receive new information in a Bayesian model) and his knowledge, attitudes and behaviors should be least unaffected. On the other hand, individuals that a priori thought their peers were substantially more conservative before learning that the "social norms" information showed that they were wrong, should be more likely to converge toward the reference points at a faster rate than individuals that underestimated the reference point by a smaller amount. To measure this potential converging effect, the baseline and endline surveys collected several indicators for each of the statistical graphs.

The second reference group was friends of study participants that were directly invited by them. Based on an individual-level randomization arm, half of treatment respondents were given extra tickets to bring up to two friends (T3). Unlike T2, the reference group was now chosen by the respondents themselves. Thus, it was more likely this second reference group was composed of a network of people that respondents considered as "significant others" (Bandura 2004) in their HIV attitudes formation.

While results of the mentioned treatment arms are reported in the main paper of the Shuga trial (see Appendix 2), the empirical papers of this thesis merge the three treatment arms¹⁹. This was required because of the nature of the moderation and mediation

¹⁹ The Shuga trial also included a third experimental arm (T4) not shown in the evaluation design table. This arm measured program spillovers on friends that were not directly or indirectly invited to the Shuga community screenings. The research team asked baseline respondents to name two friends with whom they

questions and the need to increase statistical power given the use of smaller samples.

Summary of the main study results (Banerjee, La Ferrara and Orozco 2018). The main paper (n=4,986, panel) found that Shuga had small but statistically significant effects on HIV knowledge (0.13 standard deviations, $p<0.01$) and attitudes towards HIV-positive people (0.08 standard deviations, $p<0.01$). The paper found evidence that MTV Shuga increased safer sexual practices, with percent increases often being substantial as outcomes are rare. The treatment group was twice as likely to get tested for HIV for an objective measure collected 8 to 10 months after program exposure (Control=3.3, Treatment=6.4 percent, $p<0.01$); and to report a 13 percent reduction in the number of concurrent sexual partners (C=1.32, T=1.15, $p=0.04$). In contrast, the intervention did not induce greater condom use or reduce inter-generational sex.

The key contribution of the main Shuga paper was to experimentally test the social mechanisms that potentially promoted behavioral outcomes among viewers (T2 and T3) and their friends (T4). That paper found that the increase in safer sexual practices was not mediated by conformity to social norms. Study participants in T2 did not show different effects; and there was no evidence that participants' beliefs converged towards the social reference points described in the smart graphs. In addition, watching Shuga with a friend (T3) had no differential effects. The fact that the individual channel seems to have prevailed in the Shuga trial does not imply that social norms are unimportant in general. Exposure to T1 may have already conveyed key social norms information, so T2 may have

regularly spoke. The research team randomly surveyed one from the pair. By comparing the answers of the friends of the treatment and control participants, the trial was able to test if information and values learnt through Shuga spilled over to people who had not watched it. T4 is a different sample and therefore is excluded from the moderators and mediators papers, which used the sample of study participants that were invited to the community screenings.

added little new information. Or, the artificially-induced reference groups (i.e., youth from nearby communities in T2 or the person the main participant decided to bring to the community screenings in T3) were not sufficiently influential in changing participants' beliefs and behaviors around the private behavior of sex.

On the other hand, the main paper found significant spillover effects on untreated friends' knowledge about HIV transmission, suggesting people who watched Shuga conveyed its messages to their friends. The spillover effect is driven by friends of the opposite sex, possibly proxying for respondents' boyfriends or girlfriends. However, the main paper did not find spillover effects on friends' attitudes and behaviors, suggesting that direct exposure to the show may be needed to trigger change at a deeper level.

The contributions of the DPhil research do not overlap with the analysis of the main paper. The latter did not study program impacts on high-risk individuals or the moderator variables of religion and television-watching habits. The main paper did not study the mediated effects of psychosocial and entertainment-education variables on HIV testing, for neither the short- or the medium-term measures.

SAMPLING AND DATA COLLECTION STRATEGY

Post-hoc Nature of the Thesis Studies

In randomized experiments, the random assignment of units to treatment groups justifies the causal interpretation of results in a subgroup analysis. The main condition is that the subgroup needs to be defined with baseline covariates or time-invariant characteristics (e.g., gender) as the exogeneity of treatment assignment remains not associated with

subgroup outcomes (Glennister and Takavarasha 2013; Yusuf et al. 1991)²⁰. However, the causal interpretation of results can be threatened by pre-treatment imbalances given the smaller sample. If there are systematic differences in baseline outcomes between the experimental groups, point estimates of program effects will be biased. This possibility can be ruled out by assessing pre-treatment balance between treatment and control groups with baseline data (i.e., balance tables).

The main concern about subgroup analysis is related not to confounding issues; rather, with standard errors not being corrected for multiple hypothesis testing and selectively reporting significant results (Young and Karr 2011; Westfall and Young 1993). In other words, researchers can perform many separate analyses for specific subgroups in the hope that “something will turn up” that has a p-value under 0.05²¹. Data-driven strategies have a high probability of finding a false positive. For example, if the null hypothesis is true for each of 10 independent tests for interaction at the 0.05 significance level, the chance of at least one false positive result exceeds 40 percent (Wang et al. 2007).

Although limiting the number of hypotheses with clear theoretical importance is commonly argued by researchers conducting post-hoc studies (e.g., Markham et al. 2012; Heffron et al. 2017; Molina et al. 2018), *post-hoc* studies remain greatly controversial (Wang et al. 2007). The lack of a detailed pre-analysis plan prevents researchers externally validating their decision in the selection of reported subgroups and outcomes (Chan 2005; Gamble et al. 2017). Reporting guidelines such as the Guidelines for the Content of

²⁰ What Yusuf et al. (1991) described as proper subgroups. In contrast with improper subgroups, which are demarcated by post-randomization events or measures.

²¹ On the other hand, if the tested subgroups were considered in the study design (i.e., the broader sample was stratified for the subgroup, and its subsample sufficiently powered), concerns of p-hacking and reporting bias are not warranted.

Statistical Analysis Plans in Clinical Trials (Gamble et al. 2017) require making the exploratory nature of these types of studies evident (e.g., mentioning in the title and abstract the post-hoc nature of the study).

Sample Size for the Broader Shuga Trial

Power calculations help determine the sample size of an impact evaluation (Ellis 2010). Whereas a sample that is too small means that researchers will not be able to detect a statistically significant effect, a sample size that is too large can be a waste of limited resources. Statistical power is defined as the ability of an evaluation study to detect a program's impact if there really is one. Statistical power is usually set at 80 percent, meaning that were the experiment implemented 100 times over, the evaluation would be able to detect an impact in 80 of those cases if the program's impact indeed existed.

The Shuga trial was powered at the conventional 80 percent. Power calculations were tailored to the cluster-design and to detect program impacts for females and males (i.e., the sample was powered to measure the average treatment effects by gender). Using standard deviations and intra-cluster correlations (ICC) from the 2008 NDHS microdata²²; assuming attrition rates at follow-up of 30 percent; and performing two-sided tests, the trial initially estimated 64 respondents (32 per gender) in each of the 80 study locations or a total of 5,120 individuals to measure minimum detectable effects of 0.15-0.20 standard deviations. The trial's effective panel, where we have observations for an individual both at baseline and follow-up, was 4,986. The inclusion of baseline values of the HIV outcomes of interest in the final analysis, reduced the minimum detectable effects

²² For the binary outcomes of condom use in the last sexual intercourse, ever been tested for HIV, and experienced a STI in the last year.

to 0.09-0.14 standard deviations. The effect-size literature considers these effect sizes to be small (Cohen 1988; Sawilowsky 2009).

Survey Planning and Quality Controls

The doctoral candidate led the trainings for enumerators and clinicians; and led the piloting of the instrument and of the intervention. Enumerator trainings focused on the special specifications for impact evaluation research (Iarossi 2006), especially its differences with respect to prevalence studies. For example, the need to individually identify study participants to construct panel data (and the additional measures to ensure privacy and confidentiality of respondents' information); or the importance of not discussing the potential treatment assignment and program implementation with study participants. Survey work was guided by field manuals developed by the candidate and the research team (see Appendix 10). The manuals were used to guide interviews, data collection procedures and to interpret questions in the survey instruments. The roles and responsibilities of the enumerators and team leaders were detailed in the field manual. A quality control plan based on back-check questionnaires was developed. This plan involved survey supervisors conducting follow-up interviews with randomly selected respondents selected by the research team. To monitor data quality, back-check and original responses were compared on a weekly basis. A retraining session was organized two weeks into the survey to address identified issues and to demonstrate to enumerators that their work was being monitored. Final error rates were low and within acceptable standards for high quality data collection (White 2016): 2.9 percent for variables checking that the interview was administered to the right person (e.g., age, education) and 6.6

percent for variables assessing how well surveyors administered the survey (e.g., relationship status, categorization questions, sensitive questions).

Chlamydia biomarkers were not collected at baseline due to funding constraints and to the lack of STI prevalence information for the study population. Trials that measure rare or infrequent outcomes require much larger sample sizes, as power calculations are determined by the percentage point difference between the prevalence rates of the treatment and the control groups (World Bank 2011). To address this information issue, the Shuga trial conducted a STI study pilot between the baseline and the eight-month follow-up data collection. This study pilot collected biomarkers for chlamydia, Gonorrhea, Syphilis, Trichomoniasis and Herpes Simplex Type II Virus. The sample consisted of 500 individuals (half females) between the ages of 18 and 25. The pilot found that chlamydia was the most prevalent STI, with a prevalence rate of 4.9 percent.

Recruitment Strategy

To recruit study participants, the Shuga trial employed a three-step approach. First, a listing exercise was conducted in the 80 study locations (n=17,837). The listing enabled the survey team to obtain demographic information of the households and eligible participants, which in turn formed the sampling frame for the baseline survey. Questionnaires were administered to collect information about young people between 18 and 25 years residing in the households visited.

Second, the survey team randomly selected one person in each household to be invited to a film community screening (screening zero). The survey teams returned to the listed households two to three days after the listing to invite selected participants for the community film screening. Each enumerator was provided with a list of possible invitees.

Only one young person in each of household visited was invited to attend screening zero.

The invitation cards were numbered using an automated numbering device.

Third, those who showed up at screening zero were visited at home, where they were invited to participate in the study and, if they agreed, administered the baseline survey.

The treatment groups subsequently were invited to two community screenings of MTV Shuga, with screenings being one week apart (each screening showed four of the eight episodes that make up the third season). Similarly, the controls were invited to screenings of the placebo television series Gidi Up.

Data Collection

The survey instruments were drafted by the research team based on the psychosocial and entertainment-education constructs shown in the theory of change (Figure 1).

The instruments included psychosocial and behavioral outcomes of interest, as well as important covariates and sources of heterogeneity (e.g., respondents' religion or educational level). Prior to piloting the survey instrument in the field, the instrument received feedback from the government, the implementing survey firm and NGO, and other HIV researchers. The survey instrument was designed to collect individual-level panel data, a prerequisite for studying moderator effects for time-varying outcomes (e.g., effects for individuals that engaged in risky sexual behaviors at baseline). The baseline and follow-up questionnaires included modules on respondents' demographics; knowledge about HIV transmission, attitudes towards HIV-positive people, attitudes towards risky sexual behavior, and self-reported sexual behaviors²³. The follow-up survey was

²³ The instruments also included questions regarding gender-based violence. The candidate's research outputs on gender-based violence outcomes, to be completed in 2019, are outside the DPhil research.

conducted 8 months after the intervention (it was originally planned to be conducted at six months, though contract and implementation delays moved its timing to 8 months). To improve statistical power and the precision of treatment estimates (i.e., reduce their standard errors through the inclusion of baseline values of the outcomes of interest), the follow-up instrument included the same outcome items collected at baseline. To assure comparability in responses, the follow-up survey used the same six-month timeframe of the baseline questionnaire. With the exception of the objective measures collected during the temporary clinics, all outcome measures were collected both at baseline and in the follow-up survey.

Survey items were mainly derived from the 2008 Nigeria Demographic and Health Survey, which has been validated with similar urban youth populations in Nigeria (Fatusi and Blum 2008; Odu and Foluso 2008). Example items from the Nigeria DHS instrument include the listing of ways to contract HIV, awareness of antiretroviral therapy, and respondents' willingness to disclose the HIV-positive status of a family member. The survey instrument also included items used for youth HIV-prevention research in other Sub-Saharan countries (Dupas et al. 2012, Cho et al. 2013, and Graff-Zivin et al. 2013), and a few items were specifically developed based on the scripts of MTV Shuga (e.g., how many months after the first HIV test should a person go for a second test, would you allow a HIV-positive boy to play soccer). Because most questions were derived from standardized HIV questionnaires, and because the show-specific questions represented a small fraction of the overall messages presented by the three-hour season, the doctoral candidate believes that concerns that the survey instruments were narrowly defined around the television program are not warranted. "Teaching to the test" (i.e., where the curriculum or program is heavily focused on preparing participants for a test) was not possible. The intervention

was designed independently by the MTV Foundation and before the survey instrument was designed by the doctoral candidate. Concerns that survey items for knowledge and attitudes were over-aligned to program messages at the expense of capturing these high-level constructs more accurately cannot be discarded, though these remained few in number. Moreover, although the Shuga trial took measures to minimize priming responses in the treatment group (e.g., survey firm was different than the implementing NGO, program questions were delegated at the end of the follow-up survey), a valid concern is that survey items that explicitly referred to show storylines (e.g., should an HIV+ child be allowed to play soccer) could have made salient the objectives of the research.

A series of survey pilots were conducted before the baseline to increase response variation and validate surveys items with the study youth population (Presser et al. 2004). A team of enumerators interviewed approximately 50 youth in a two-week period. The instrument was revised according to the pilot findings. Questions were revised in cases where over 10 percent of respondents chose “other” or when responses suggested under (over) reporting²⁴. The survey instruments were designed, and the enumerators were trained with the goal to minimize social desirability bias (Brener et al. 2003; Minnis et al. 2009) following best practices in survey implementation (Iarossi 2006). For example, surveys were conducted in private spaces and enumerators were the same gender as the respondent.

²⁴ For example, an unusual high percentage of females (>50%) responded not having started sexual activity when they were directly asked this question. To reduce this potential source of social desirability bias, the question was reframed to: *when was the last time you had sex?* The response options included “the individual had yet to start sexual activity”. This new framing increased the number of positive responses, increasing the statistical power of this potential outcome. Given the random assignment of the treatment, neither framing would have biased the impact estimator.

Table 3 shows the trial’s timeline for the average study participant. After the eight-month follow-up survey, respondents were provided with phone-credit incentives to return two weeks later to the 80 community centers. These centers now served as “temporary clinics”, which helped collect objective measures for condom demand, HIV testing, and chlamydia infections. For the condom demand measure, clinic attendees were randomly offered the choice of receiving condom packs or their financial equivalent. It was theorized that the treatment group would be more likely to choose condom packs compared to the control group. The doctoral candidate carried out a similar condom-demand experiment during a follow-up data collection in Malawi (Cho et al. 2013), making revisions for the Shuga trial²⁵.

Table 3. Timeline for the average study participant

Months						
0	1	2 - 7	8	9	10-11	12
<u>Enrollment</u> -Eligibility screened and informed consent -Baseline survey (Oct 2014)	<u>Interventions</u> -Community screenings treatment and control locations (Nov 2014)		<u>Assessment 1</u> -Follow-up survey - Individuals given information about HIV testing centers (Jul 2015)	-Temporary clinics conducted (Aug 2015)		<u>Assessment 2</u> -Research team collects notes at HIV clinics (Oct 2015)

To collect the objective HIV testing measure, the doctoral candidate designed an innovative approach where clinic attendees were provided with information notes with

²⁵ The goal of this game was to proxy demand for condoms as the “almost free” price of a condom pack guarantees that the vast majority of participants can choose the condom packs if they value their health benefits. The Malawi trial found no effects on condom use in both self-reported condom use and the objective condom demand game. The Shuga trial tried to improve the validity of the condom game in two ways. First, the condom game was conducted at the temporary clinics and not as part of the follow-up survey, providing greater privacy for choosing condoms. Second, the condom game offered a larger number of condoms and financial incentives, which may have potentially increased the cognitive efforts invested by participants in their decision-making process (Westbrook & Braver 2015).

the address of nearby HIV testing centers. The research team visited all HIV testing centers two months after the temporary clinics were conducted. The information notes included a unique identification number which allowed tracking study participants that went to get tested for HIV (not their test results). Similarly, it was hypothesized that the treatment group would be more likely to get tested than the control group. Unfortunately, it was not possible to learn when exactly in this two-month window study participants visited the HIV testing clinics.

Finally, clinic attendees were offered the opportunity to test for chlamydia, based on the “three C’s” - informed consent, counselling and confidentiality (World Health Organization, WHO 2013). After a brief counselling session²⁶, participants who consented to testing for chlamydia were provided with sterile bottles to collect mid-stream urine specimens by themselves. Participants’ urine specimens were tested for the presence of chlamydia infection. Testing methods included a combination of routine tests and rapid, serological tests. For quality control, a second lab scientist validated the initial results and countersigned them. Clinic attendees were then invited to return two weeks later to receive *post-test* counselling and receive their test results. Medical doctors prescribed appropriate treatment for those who tested positive for chlamydia.

To improve the research team’s supervision of the data collection and of the intervention, the timing of the baseline survey and of the community screenings was staggered: in each of the 80 locations, the implementing NGO carried out the community screenings two

²⁶ Trained counsellors took participants through a pre-test counselling session that included seeking informed consent. The counsellor carried out a risk assessment, helped the participant develop a risk reduction plan, and explored potential obstacles that participants may encounter when implementing this plan.

weeks after the survey firm had collected the baseline survey. The follow-up survey and the temporary clinics followed a similar staggered strategy.

Trial flow diagram and Follow-up attrition

High and systematically different attrition rates between treatment and control groups at follow-up can potentially introduce attrition bias in experimental studies (Dumville, Torgerson and Hewitt 2006), affecting the validity of impact estimates. Figure 5 shows the trial's CONSORT flow diagram, describing the number of participants who were randomly assigned, received the intended treatment, and that were analyzed. The listing collected information from 17,837 households from the 80 study locations. At baseline, 5,166 study participants agreed to be part of the study and completed the baseline survey. The eight-month follow-up survey managed to track and interview 4,986 baseline respondents, which corresponds at follow-up to an attrition rate of 3.5 percent. Thus, outcomes measured by the eight-month follow-up survey are unlikely to be affected by attrition bias.

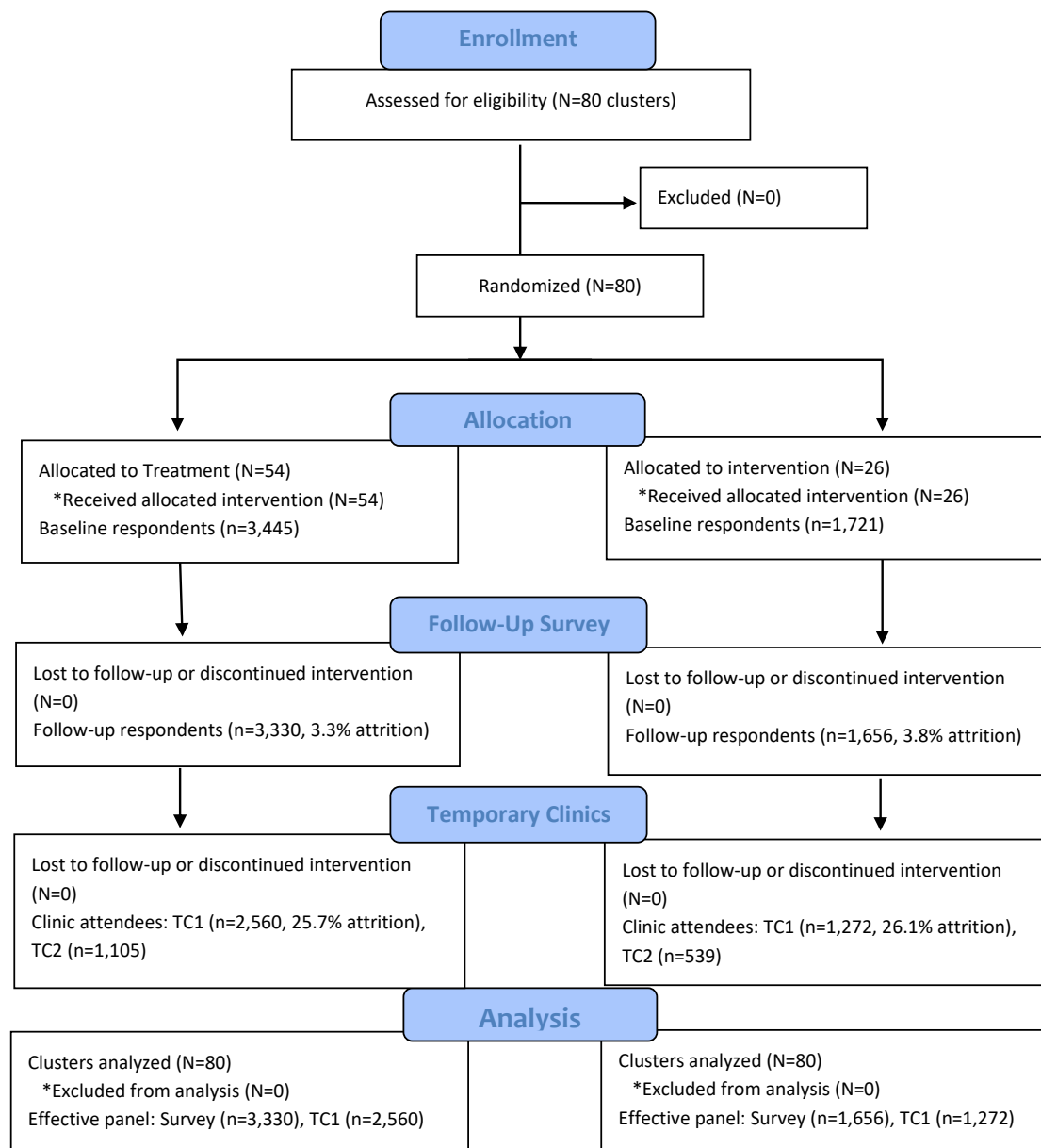
Only follow-up respondents that remained in the city where they were interviewed at baseline or that moved to one of the other six study cities were financially-incentivized to attend the temporary clinics. At the time of the follow-up, a quarter of the baseline respondents had moved outside the study cities (mostly to attend university). While the survey firm interviewed these individuals in person, these respondents were not expected to attend the temporary clinics. Out of the 3,901 respondents that remained living in the study cities, 3,832 attended the first temporary clinic. This corresponded to an attrition rate of 1.8 percent with respect to the in-study-cities sample (or an attrition rate of 25.8 percent with respect to the baseline sample, with the difference being driven by

individuals that were not expected to attend the temporary clinics). Selection bias (Heckman 1979) is also unlikely to affect the HIV objective outcomes collected through the temporary clinics. As reported in the main paper, analysis of attrition rates for the follow-up survey and for the temporary clinics indicate that missing data patterns did not vary systematically by treatment assignment. Thus, study participants can still be considered part of the same population regardless if they were retained at both post-treatment data collections (Dumville, Torgerson and Hewitt 2006; World Bank 2011).

Most relevant to this thesis, and as shown in the moderators paper (Orozco 2019), attrition and selection biases were not threats for the subsample of high-risk individuals. This evidence, together with the fact that none of the study 80 locations were dropped in the final analysis and evidence of pre-treatment balance between treated and control groups, strengthen the experimental interpretation of the subgroup analysis. However, its post-hoc nature prevents calling study results as hypothesis-testing evidence.

Finally, 1,644 individuals returned to the second temporary clinic to receive their chlamydia results and 186 clinic attendees went to a nearby HIV testing center to get tested. Because both events are study outcomes, they should not be interpreted as attrition rates.

Figure 5. Shuga CONSORT flow diagram



Data Challenges

The data collection faced three important challenges, partly addressed with the following mitigation measures:

(1) Relocation of participants. Internal roving teams by the survey firm were deployed to track and interview eligible participants wherever they were in the country.

(2) Wrong contact details of participants due to the lack of clear addresses. Using Google Earth maps showing GPS locations of respondents, the follow-up traced participants to the houses where they were during the baseline survey.

(3) Refusal to participate. As noted above, phone call cards equivalent to US\$2.5 and US\$7.5 were provided to complete the surveys and attend health camps respectively. These amounts were determined through a pilot that tested different payment levels to maximize the take-up of the invitations. Because the outcome of interest (e.g., going to get tested) was not incentivized, and because both experimental groups were offered the same amount, these relatively large incentives should not affect impact estimates (the control nets out the effect for the treatment incentives) and the generalizability of results (objective outcomes were not incentivized).

ETHICAL CONSIDERATIONS

Informed Consent

Survey enumerators from the survey firm (Hanovia Medical Limited) provided the participants with the study's information and consent forms. Enumerators assessed if the study participant was voluntarily and knowingly giving informed consent. Enumerators

were instructed to spend as much time as necessary to listen to respondents' questions and concerns about their participation.

After study participants signed the consent form, they were reminded that they were free to stop answering the survey questions at any time. Informed consent was also sought at the temporary clinics by health personnel from the Department of Medical Microbiology and Parasitology at the University College Hospital in Ibadan. The consent forms for the collection of surveys and STI biomarkers can be found in Appendix 5.

Privacy, Data Storage and Security

The trial's data used personal identifiers to maintain confidentiality of participants' information. During the baseline survey, participants were assigned a unique numeric study identifier. The follow-up survey used this same number to create the individual-level panel dataset. A master list linking the names of the individuals to the numeric study identifiers were kept in an encrypted file in a password-protected computer and were kept secure by the study Principal Investigators. Similar protocols were followed to protect the identity of temporary clinic attendees and prevent disclosure of their chlamydia results.

Individual data is aggregated for the final analysis, so individuals cannot be identified in any publication. De-identified data and study documents will become publicly available through the World Bank Microdata Library after the publication of the Shuga main paper. As the main output of a World Bank project, the main paper will go through a series of protocols and clearances by senior management of this institution.

Potential Risks of Participation

The trial envisioned two potential risks of participation. First, study participants may have risked social harm if confidentiality of the survey information and of the chlamydia results were breached. The research team ensured that the interviews and STI biomarkers collection were conducted in a private location and that staff was trained to maintain confidentiality. No breaches of confidentiality were reported. Second, study participants who tested positive for HIV may have risked psychological harm if they lacked access to appropriate standard of care for HIV, which includes antiretroviral treatment in Nigeria. To facilitate referral to treatment and to minimize psychological harm, the information notes provided at the temporary clinics only included addresses of accredited HIV testing centers that offered or were linked with facilities that offered this standard of care for HIV in Nigeria. As noted above, clinic attendees that tested positive for chlamydia were offered treatment during the second temporary clinic. No post-treatment issues were reported to the research team.

Ethical Approval

Ethics approval for the baseline survey, STI pilot, and the follow-up survey and biomarkers data collection was secured from the National Health Research Ethics Committee of Nigeria (NHREC/01/01/2007-14/07/2014) and the ethics committees of Massachusetts Institute of Technology (MIT, 1401006127) and Bocconi University (170314). Appendix 6 includes the initial approval letters from the National Health Research Ethics Committee of Nigeria, MIT and Bocconi University. Oxford's OXTREC (564-15) provided ethical clearance for the follow-up survey and biomarkers data collection. The OxtREC application and approval letters are respectively found in Appendixes 7 and 8.

MEASURES

Outcome Selection for Thesis

The moderators and mediators papers jointly used eight intermediate and final outcome measures, which can be categorized into three types of variables: *Psychosocial* (HIV knowledge and attitudes towards HIV-positive people), *Entertainment-Education* (transportation) and *Behavioral outcome variables* (sexual concurrency, unprotected sex, inter-generational sex, HIV testing and chlamydia biomarkers). The Shuga trial collected data on behavioral intentions²⁷, though the thesis does not use them to focus on realized behavior change. The baseline and the follow-up instruments can be found in Appendix 9A and 9B, respectively.

A variety of self-reported measures were examined during the survey work. The majority of scales had been used in previous studies with Nigerian urban youth, though reliability measures for their use were not found. Most of the psychosocial items were derived from the instrument of the 2008 Nigeria Demographic Health Surveys. This population-level survey has been validated with youth urban populations in Nigeria (Fatusi and Blum 2008; Odu and Foluso 2008), repeatedly being administered for this population since 1990. A series of survey pilots were conducted prior to the baseline to increase the response variation and validate the psychosocial variables in the study population (Presser et al. 2004). The instrument items had minor changes after the piloting of the instrument, confirming response variation and high-reliability (Sullivan 2011) of the different questionnaire items. The validation of the entertainment-education items was not

²⁷ Behavioral intentions were mostly hypothetical questions that were only relevant for individuals that had yet to started sexual activity. For example, these individuals were asked if they were planning to use a condom the first time they had sex.

conducted for financial and time constraints, as their validation required conducting screenings. The interviewer's instruction manual for the baseline can be found in Appendix 10.

Researchers often face a dilemma when choosing items that best fit their research project. Considering that most psychology studies that intend to measure a psychosocial construct will use different instruments or modules (to adjust to its intervention and population of interest), it seems that "generalizing" about a high-level psychosocial construct is often not relevant. Although there is consensus that psychology research should maximize the prediction of their instruments by using validated instruments, *there is less consensus regarding the predicted match in terms of content domain or the level of the construct being targeted by the intervention or measured by the study* (Rosip and Hall 2004). This partly explains why there are no clear guidelines for choosing specific instruments for studying the effectiveness of different interventions and their outcomes in psychology (Hall, Andrzejewski and Yopchick, 2009).

The analysis is based on individual survey items and constructed indexes. Items from four of the eight outcome measures used in Orozco (2019) and Orozco and Cluver (2019) were aggregated into indexes. The use of indexes (1) increased statistical power by reducing errors due to random variation through aggregation (Jackson 2003) and (2) helped reduce Type-I error inflation²⁸ due to multiple-hypothesis testing and concerns of "p-hacking" of significant results (Young and Karr 2011). The combined 21 indicators for HIV knowledge and attitudes were aggregated into two outcome indexes. The indexes were coded in a

²⁸ In the context of statistical hypothesis testing, a Type-I error or "false positive" is the rejection of a true null hypothesis (i.e., to falsely infer the existence of an effect that does not exist). Type-II error is failing to reject a false null hypothesis or a "false negative" (i.e., to falsely infer the absence of the effect when this exists).

way where answers consistent with Shuga's objectives translated into higher values. For their construction, principal component analysis for a family of outcomes was used (Anderson 1963, Jackson 2003). Principal component analysis is used to reduce the dimension of the variable space. Through an orthogonal transformation, it approximates interrelated objects by extracting their data differences into several independent principal components. The principal components are then ordered so that the first one has the largest possible variance, with the method selecting components that best represent the correlated variables.

Cronbach's alphas (Cronbach 1951) measure reliability or internal consistency, or the average inter-correlation among a set of items. It assesses how well an aggregate-test measures what it intends to. A commonly accepted rule is to consider Cronbach alphas above 0.70 acceptable (Kline 2000; George and Mallery 2003). Three of the four indexes concentrated around this threshold. The exception was the HIV Attitudes index, which was substantially below this threshold ($\alpha=0.47$). This low value suggests that the aggregated index included sub-concepts that were not narrowly related (Streiner 2003). This was partly the goal for the attitudes (and other indexes), as they intended to be comprehensive measures. Concretely, the HIV Attitudes index included three related sub-concepts discussed in the three-hour drama: (1) willingness to reveal HIV status of a family member or of oneself; (2) stigmatizing attitudes towards HIV-positive people (i.e., would buy vegetables from or allow playing an HIV-positive person); and (3) attitudes towards HIV testing (i.e., getting tested means that someone has been "sleeping around"). Although the main reason of aggregating attitudes items (in a post-hoc manner) was to reduce reporting bias concerns of "selectively" reporting items, the aggregated items are also conceptually closely related. For example, an individual that holds negative attitudes

towards HIV-positive people (e.g., considers them “libertines”) will probably have similar negative attitudes towards those who get tested (e.g., well-behaved people don’t need to get tested) and would be more likely to hide their HIV-positive status (e.g., a consequence of perceived moral failure).

This comprehensive measure contrasts with mediation studies where the mediating variable is not an index; or included items are fewer and highly related with each other and with the behavioral outcome. Thus, the indexes used in this thesis may not be reliable measures for research projects studying more focused interventions and narrower constructs. In line with Rosip and Hall (2004) and Hall, Andrzejewski and Yopchick (2009), the candidate believes there is a methodological trade-off to be considered between measuring higher-level constructs (e.g., general attitudes towards HIV-people) with greater noise or measuring lower-level constructs (e.g., willingness to reveal one’s status that can be linked to the outcome of actual disclosure) with less noise. The former case may be more appropriate for interventions with a comprehensive list of messages, like entertainment-education.

Measurement error²⁹ of the intermediate outcomes are likely to adversely affect thesis results. The use of variables with low internal consistency makes less likely to observe statistically significant effects in the moderation analysis due to the larger variance of the outcome of interest (Moonseong, Namhee and Myles 2015). In most cases, this increased variance will also affect the statistical significance of the mediating estimators (Fritz, Kenny and MacKinnon 2016). Moreover, measurement error may also bias the mediating estimator, with the direction of the bias depending on the relationship between the

²⁹ Irrelevant factors that influence a score of a variable besides the theoretical construct of interest.

indirect and the direct-effect variables (Fritz, Kenny and MacKinnon 2016). When both have the same sign, the mediating estimator is inflated. When they have opposite signs, it is underestimated. In the thesis mediation study, the direct and indirect effects have the same sign, hence potential measurement error of the mediator variables is likely to overestimate mediating estimates (Le Cessie et al. 2012; Fritz, Kenny and MacKinnon 2016).

Removing the less-related items did not qualitatively change the base value of $\alpha=0.47$ for the Attitudes index. This result is not surprising partly because Cronbach values tend to decline with fewer items (Le Cessie et al. 2012). Cronbach alphas were under 0.40 for the two-item sub-concepts. Thus, in addition to highlighting the limitations that measurement error may have in the thesis empirical papers, both papers also note that these indexes may not be reliable measures for future research projects. Interestingly, the low α of the HIV Attitudes index was partly counterbalanced by its larger and significant effects in both the moderators and mediator's papers with respect to the Knowledge index, which had an α of 0.72.

Psychosocial Variables

Knowledge about HIV transmission and benefits of HIV testing index. This index included 14 items (Cronbach $\alpha=0.72$) and aimed to capture a respondent's awareness and knowledge about HIV transmission mechanisms; the existence of antiretroviral drugs; and knowledge about the adequate timing for HIV testing. Response options were binary (True, False), multiple choice, and open questions. The knowledge module first asked respondents to provide as many answers to the following two open questions: (i) How can someone get HIV? and (ii) What can a person do to reduce their chances of getting HIV?

Potential answers also included common misconceptions about HIV/AIDS, such as contracting the virus through mosquito bites. These questions were then followed by three closed questions about mother to child HIV transmission that could be answered with (Yes, No, Don't know) answers: Can HIV be transmitted from a mother to her baby: (iii) During pregnancy? (iv) During delivery? (v) By breastfeeding? and (vi) Are there any special drugs that a doctor or a nurse can give to a woman infected with HIV to reduce the risk of transmission to the baby? (vii) Have you heard about special drugs that people infected with HIV can get from a doctor or a nurse to help them live longer? For verification purposes, individuals that responded "Yes" to item (vii) were asked (viii) if they remembered how these drugs were called. Shuga also discussed the importance of waiting three months after potential exposure to HIV to get tested. To measure the impact of this message, the following four items were developed for this study: (ix) Suppose someone thinks that [he/she] may have been exposed to HIV, for example through sex or through a contaminated blade. Suppose this person immediately goes for an HIV test and it is negative. Can this person be sure that [he/she] has NOT been infected with HIV? (Yes, No, Don't know) (x) Why can't he/she be sure? (xi) How many months after the first test should this person go for a second test? (Number of months).

Stigma and attitudes towards HIV-positive people index. This variable included 7 items (Cronbach alpha=0.47) and intended to capture respondents' inclination to potentially reveal their status, allow HIV-positive people to interact with the community, and not hold negative judgements towards these people. Response options were binary (Yes, No) and were based on a 4-point scale (Strongly agree, Agree, Disagree, Strongly disagree). The survey items asked respondents': Do you think a young boy with HIV should be allowed to play football with other boys? (Yes, No, Don't know/ not sure/depends).

The instrument then asked stigma items (with similar response formats) that are included in the Nigeria DHS population surveys. Specifically, respondents' willingness to disclose the HIV status of a family member; willingness to purchase fresh vegetables from a shopkeeper with HIV/AIDS; believes people that get tested is proof they have been sleeping around; believes people with HIV should be blamed for bringing the disease into the community; and believes HIV/AIDS as a punishment for sleeping around.

Entertainment-education Variable

In the last module of the follow-up survey, the instrument asked respondents in treatment and control groups if they recalled the title and the main themes of the television dramas each group watched. These questions were then followed by items that aimed to measure program impacts on the potential entertainment-education mediators of transportation.

Transportation index. This variable included 12 items (Cronbach alpha=0.78) and aimed to measure individuals' levels of transportation and emotional involvement when watching the treatment and placebo programs. The items used for the Transportation index were adapted from Green and Brock's transportation scale (2000). Response options were based on a 5-points scale (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree). Example items included asking individuals if they were emotionally affected by the television program and if their mind was wandering during the community film screenings.

Behavioral Variables

Prior to starting the sexual behavior module, enumerators reassured individuals that their answers were completely confidential and that they were free to skip specific questions

at any time. This module began with the gateway question “When was the last time you had sex?” (has had sex before, has never had sex, refused to answer). For those who answered affirmatively, this question was followed by the item “Have you had sex during the last 6 months? (Yes, No, No Answer). The survey instrument asked sexual behavior questions for each of the respondents’ last three sexual partners (e.g., did you use a condom with partner X1, X2 and X3).

HIV Testing. This outcome was studied through two indicators, which measure the period before the follow-up survey (“short-term”) and after it (“medium-term”). Follow-up respondents that reported having at least once gotten tested for HIV were asked when the last time was they got tested. Response options included: less than six months ago, 6-12 months ago, 12-13 months ago, and two or more years ago. A binary variable was constructed for individuals that reported getting tested in the last six months. The “medium-term” indicator was an objective behavior measure. Clinic attendees were provided with information notes with the address of the closest HIV testing centers. The information note had an individual identification code and was directly collected by the research team at the HIV testing centers within two months of the follow-up survey. Following Baird et al. (2012), study participants were not offered HIV testing at baseline as this would have likely changed their sexual behaviors and reduced the number of individuals testing for HIV at follow-up, a primary study outcome.

Chlamydia status. Clinic attendees were asked to provide urine samples to test for chlamydia through a combination of routine and rapid tests. Rapid tests included a qualitative, lateral flow immunoassay to detect the chlamydia antigen. STI biomarkers were only collected at follow-up.

Sexual concurrency. Concurrent sexual partners is a relatively new concept that was added for the first time to the Nigeria Demographic and Health Survey in 2013. This outcome is defined as the percentage of respondents who had two (or more) sexual partners concurrently. This high-risk sexual behavior was measured through two-self-reported questions, which focused on two different time periods. Survey respondents were first asked if they had engaged in this practice in the last six months (Yes/No, a binary outcome). Respondents were then asked the number of sexual partners they currently had at the moment of the follow-up data collection (quantitative outcome).

Unprotected sex. This outcome was measured in two ways. First, survey respondents were asked if they used a condom in their last sexual encounter. The question was repeated for the last sexual encounter with their primary and non-primary partners (the latter required the respondent reporting at least two sexual partners). Second, an objective measure for condom demand was collected at the first temporary clinic. Attendees were offered either condoms packs or their financial equivalent to clinic attendees. It was theorized the treatment group would be more likely to choose the condoms packs compared to the control group. This “artificial” objective measure for condom-demand should be interpreted as part of a series of condom-use measures collected by the survey work. Thus, if the intervention was effective in promoting condom use, the point estimates of objective and self-reported outcomes should move in the same direction.

Inter-generational sex. This behavior outcome was measured by asking survey respondents the age of their last three sexual partners during the past six months. An inter-generational binary indicator was constructed for individuals that reported an age

difference of 10 years or more. This age-difference gap follows the definition commonly used in HIV research (Balkus et al. 2015; Schaefer et al. 2017).

Construction of Variables

The construction of moderator variables in the moderators paper (Orozco, 2019) was conducted in the following way: a binary variable (1=high-risk individual) was constructed for participants who reported at baseline at least one of the four HIV-risk factors (namely, had unprotected sex in their last sexual encounter, engaged in sexual concurrency, engaged in inter-generational sex or reported having symptoms of an STI). The effect-modification analysis in Paper 2 was restricted to religion and television-watching habits for theoretical reasons³⁰. The religion moderator variable was binary (Muslim=1, otherwise=0). Practically all non-Muslims were Christians in our sample as both categories covered over 99.5 percent of study participants. Responses were categorical for television consumption levels (Do you watch television almost every day, at least once a week, less than once a week or not at all?). A binary variable was created for individuals that reported watching television every day to create daily and non-daily viewers arbitrarily. The choice for daily television-viewers divided the sample near its median (45 percent of respondents reported watching television every day). With the goal to reduce the loss of statistical power, median splits are commonly done in experiment studies (Iacobucci et al. 2015).

The mediators paper did not require constructing additional variables to those already reported for the moderators paper.

³⁰ Because there were no clear theoretical reasons why program impacts may have differed for other potential moderators (e.g., education or income level), and to minimize concerns of data mining in terms of selecting subgroups where differences are significant, the effect-modification analysis was restricted to religion and television-viewership levels (i.e., the candidate did not test other moderator variables).

STATISTICAL ANALYSIS

Technical Considerations for the Thesis Papers

Because the moderation analysis is based on a smaller sample for which the Shuga trial was not designed, the high-risk subsample may potentially (i) be imbalanced or affected by selective attrition affecting the internal validity of findings (i.e., biased point-estimates) or (ii) not be sufficiently powered to statistically detect effect heterogeneity (i.e., given the smaller sample, higher minimum detectable effects are required to observe similar statistical results), increasing the likelihood of Type II error (i.e., effect is true but cannot be statistically detected). The first issue was ruled out through baseline and attrition analysis.

The power limitations remain, however, an inherent limitation of post-hoc moderation studies. The decrease of chlamydia infections presents an example. While the Shuga main paper found a statistically significant decrease for this outcome on women, the moderation paper did not, despite the effect size being slightly lower in absolute terms (1.3 versus 1.7 percentage points). Using the same power and significance level assumptions and observed inter-cluster correlation for this outcome, the sample size required to statistically detect the observed effect in the moderation analysis was 7,720 women³¹. The number of women in the moderation analysis was only 772. However, the null hypothesis of no effects cannot be ruled out for the high-risk population³².

³¹ In contrast, the slightly larger effect observed for the larger study required a total of 2,240 women. This sample was sufficient to find significant effects at the 10 percent level for the average study participant.

³² A “mismatch” problem (Institute of Education Sciences, 2008) occurs when random assignment is carried out at the cluster level (e.g., community screening location) and the analysis is conducted at the individual level (e.g., HIV testing), but the correlation among study participants within the same clusters is ignored in computing the standard errors of the impact estimates. The study’s cluster design, which intended to

The analysis of Paper 2 and Three reports effect sizes, 95% confidence intervals and p-values. The significance level, a pre-chosen probability of type I error, was set at the conventional 5 percent for all three DPhil papers. The p-value is described in our analysis in terms of rejecting the null hypothesis when it is actually true (Glennerster and Takavarasha 2013). For example, a $p=0.05$ would mean that there is a 1 in 20 chance that the significant differences we observe between treatment and control outcomes are not true (i.e., falsely rejecting the null hypothesis, or in our study case, wrongly concluding there was a difference between our treatment and control groups). The smaller the p-value, the lower the chance of getting a difference as big as the one observed if the null hypothesis were true. In other words, the smaller the p-value, the stronger the evidence against the null hypothesis.

Following the common practice in public health and other social sciences, outcome differences with p-values greater than 0.05 are referred to as not statistically significant (including for the case of $0.06 < p < 0.10$). The 0.05 value is an arbitrary threshold. The moderation study reports the p-values together with the 95% confidence intervals. Interpretation of confidence intervals should not focus on whether or not it contains the null value, but on the range and potential importance of the values covered in the interval (Kirkwood and Sterne 2003). Clinically important differences can potentially inform future research projects. The thesis discusses both the clinical and statistical significance. Statistical difference does not imply a meaningful impact, a common mistake in the economics literature (Ziliak and McCloskey 2004). Large sample sizes may detect

minimize contamination between treatment arms, drives upwards the sample-size requirements. With similar assumptions, individual-level randomization required only 750 women.

statistically significant effects that are small in magnitude, and potentially not relevant for policy decision-making and program investments (Wasserstein and Lazar 2016).

The moderation and mediation analyses were planned from the beginning as part of the candidate's thesis. However, because these papers lacked a detailed pre-analysis plan, and because the analysis was not pre-specified in the (public) research protocol, concerns regarding data mining and reporting bias cannot be externally validated. As a result, the papers' titles and text highlight the post-hoc nature of the analysis following the Guidelines for the Content of Statistical Analysis Plans in Clinical Trials (Gamble et al. 2017). Research suggests that reporting bias, when statistically significant effects are more likely to be reported than non-significant effects, is among the most substantial biases affecting results from individual studies (Chan 2005).

The moderators and mediators papers did not control for family-wise error rates (FWER, Westfall and Young 1993) for four reasons. First, correcting for this type of error is less relevant for hypothesis-generating research, where causal claims are not being tested (Hartwick and Barki 1994). Second, the tested outcomes are independent constructs and as such, do not fall under one family that needs to be corrected for. On average, each of the tested outcomes had two measures. Third, the thesis limits the number of *subgroups* to those with strong policy and theoretical relevance. The moderation analysis focuses on high-risk individuals, exploiting the trial's gender-stratification to measure program effectiveness per gender. Finally, the analysis substantially reduced number of *outcomes* to those with strong policy and theoretical relevance. The number of outcomes in the moderators paper was seven (HIV testing, chlamydia status, sexual concurrency, unprotected sex, inter-generational sex, knowledge, and attitudes), all important

outcomes in the HIV literature. The mediation analysis tested one outcome in two periods (with the Introduction chapter providing a detailed reasoning behind its selection) and three potential mediators, all central constructs to the theoretical framework described in Figure 1.

Following the guidelines from the American Psychological Association (Wilkinson 1999), the moderators and mediators papers reported unstandardized measures (e.g., mean differences for binary and continuous outcomes) rather than standardized measures (e.g., standardized mean differences) given that the units of measurement for the behavioral outcomes are meaningful on a practical level (e.g., increase in percentage points of HIV testing). To facilitate interpretation of effects on HIV Knowledge and Attitudes indexes, impacts were converted to standard deviations. Index effects were divided by standard deviations derived from the control group at follow-up. An increase of two standard deviations, for example, would move someone from having average knowledge to being in the top five percent of the group.

Moderator Analysis

Moderator analyses are performed to assess whether treatment effects vary across subpopulations (Kraemer 2006). Two usual questions of interest are:

1. Did the program work for each subgroup? and
2. What is the difference in treatment effects between subgroups?

The first question, referred to in the moderators paper (Orozco 2019) as the *subgroup analysis*, measures average treatment effects by comparing mean outcomes of treated and control groups for a given subgroup. In our case, did the intervention work for high-risk youth, including high-risk females and males? Given the assumption that program effects will go in the same direction for both genders (reducing the statistical power to detect any potential difference), the subgroup analysis makes no attempts in testing if the treatment difference differed across genders (i.e., were the effects for women larger/smaller than the effects on men?).

The second question, referred to as the *effect-modification analysis*, is based on the treatment-moderator interaction term (Gerber and Green 2012) which measures the outcome difference between two subgroups. The effect-modification analysis studies the differential effects for Muslims (with respect to Christians) and the differential effects for daily television-viewers (with respect to non-daily viewers). To isolate their moderating influence from gender effects, the effect-modification analysis is separately conducted by gender. The effect-modification analysis is more likely to be powered than testing if treatment effects differed by gender, as the theoretical effects of the former are likely to have different effect slopes (e.g., daily viewers are less likely to benefit than individuals

that occasionally watch television). This in turn potentially increases the outcome gap between both groups, increasing the statistical power of difference tests.

The subgroup analysis is the primary analysis of the moderators paper (Orozco 2019). It benefits from (i) the Shuga trial being powered to detect effects by gender; and for the high-risk sample, observed (ii) pre-treatment balance and (iii) low and unsystematic attrition rates among the treatment and control groups. Given these three points, estimates of the subgroup analysis should approximate those of hypothesis-testing evidence. However, because the trial's sampling strategy was not powered for the high-risk subgroup (which represent approximately 40 percent of the sample per gender) and because the subgroup analysis is post hoc (i.e., not pre-specified), the primary analysis should be read as hypothesis-generating evidence (Spieth et al. 2016).

The effect-modification analysis serves as the secondary analysis. This analysis has stronger limitations than the subgroup analysis. The effect-modification analysis is based on a quarter of the overall sample and is more exploratory in nature.

The moderators papers uses analysis of covariance (ANCOVA). ANCOVA is a general linear model that evaluates whether the means of a dependent (outcome) variable are equal across levels of a categorical independent variable (treatment assignment), while statistically controlling for the effects of other covariates. This method decomposes the variance in the outcome variable into variance explained by the covariates, variance explained by the treatment variable, and residual variance.

Equation 1 shows the ANCOVA specification used to estimate the average treatment effects for the primary analysis.

$$(1) \quad Y_{ilc1} = \delta Y_{ilc0} + \beta T_{ilc0} + \mathbf{X}'_{ilc0} + \Phi_c + \varepsilon_{ilc1}$$

where: Y_{ilc1} indicates the outcome for individual i who lives in location l within city c during the follow-up (time 1); Y_{ilc0} is the baseline value of the outcome of interest; T_{ilc0} represents treatment indicators is a dummy equal to 1 if the individual was assigned to the Shuga treatment and 0 if assigned to the control; $\mathbf{X}'_{ilc0}$ is a vector of controls measured at baseline (for consistency in the analysis, this vector includes the same controls used in the main paper that are potentially associated with the outcome, improving the precision of treatment estimators)³³; and Φ_c are city fixed effects (7 towns). The standard errors were clustered at the screening-location level (80 clusters).

Modeling heterogeneity involves specifying in the regression how moderator variables contribute to higher or lower treatment effects. Equation 2 shows the ANCOVA specification used in the effect-modification analysis for religion and television consumption levels. The specification is the same as Equation 1, with the difference that both moderators are controlled for and the regression includes the interaction terms (Gerber and Green 2012) of the treatment with the moderators ($T_{ilc0} * M_{ilc0}$).

$$(2) \quad Y_{ilc1} = \delta Y_{ilc0} + \beta T_{ilc0} + \lambda(T_{ilc0} * M_{ilc0}) + \mathbf{X}'_{ilc0} + \Phi_c + \varepsilon_{ilc1}$$

All analyses were done by intention to treat (ITT)³⁴. ITT analysis means all study participants that were randomly allocated to treatment are included in the analysis regardless if they received the assigned treatment or not (i.e., attended the community

³³ These included: gender, respondent was Muslim, age, currently enrolled in school, number of years of formal education, spoke English as primary or secondary language, spoke Yoruba as the main language at home, was not living with parents, number of household members, dwelling where the respondent lived was owned, the father (and mother) having achieved more than secondary education, and a wealth index constructed from house characteristics and durable goods.

³⁴ Per protocol estimates would approximately be 20 percent higher considering attendance rates to community screenings was around 80 percent. ITT estimates were used to minimize concerns of selection issues.

screenings)³⁵. ITT analysis replicates better what happens in the “real world”, incorporating non-compliance or poor adherence of individuals. As a result, it provides a lower-bound estimate of program effectiveness.

Following 2010 CONSORT reporting guidelines (Moher et al. 2010), unadjusted means and standard deviations are presented for both treatment and control groups at follow up. The analysis also reports the adjusted impact estimates and their respective p-values when controlling for a series of covariates, including the baseline value of the outcome of interest. The observed HIV testing and chlamydia biomarkers were not collected at baseline, though this does not affect the experimental interpretation of these outcomes (e.g., Baird et al. 2012). Because the intervention was randomized, impact estimates could be presumed to be unbiased as both the treatment and control groups are in expectation identical (Bruhn and McKenzie 2009; Todd 2014). Furthermore, the baseline data finds balance between both groups on observable characteristics. This corroborates the experimental interpretation of STI biomarkers when only using post-intervention data. However, an important limitation of lacking baseline values for a time-variant outcome is that it prevents studying heterogeneous questions on that outcome. For example, what is the program’s average treatment effect for individuals that tested positive for a STI at baseline; and did program effects vary for this subpopulation versus study participants who tested negative at baseline?

³⁵ In contrast to per-protocol analysis, which only includes individuals who adhered to the treatment they were assigned to.

Mediation Analysis

To give sufficient time for the intervention to potentially impact the HIV testing outcome in the eight-month follow up survey, the sample in the mediation analysis was restricted to study participants that had already started sexual activity at baseline (i.e., sexually inactive needed to start a relationship and subsequently get tested for HIV within a short time window. Thus, their inclusion would have likely affected the study's power). The mediation analysis aims to understand the cognitive and non-cognitive mediating variables that may explain the increase in HIV testing in the short- and medium-term³⁶. See Figure 6 (a variation of Figure 1 that focuses on the subset of tested mediators under a mediation framework). The solid lines describe the mediating relationships being tested under the mediation framework of indirect and direct effects (MacKinnon et al. 2002)³⁷. *The study's theoretical model assumes that program impacts on the three mediating constructs happened after being exposed to the intervention and before the behavior impact.* As a result, these impacts are assumed to precede in time impacts on HIV testing, fulfilling the time-sequence requirement of mediation frameworks (i.e., mediators subsequently affecting the final outcomes, Baron and Kenny 1986)³⁸.

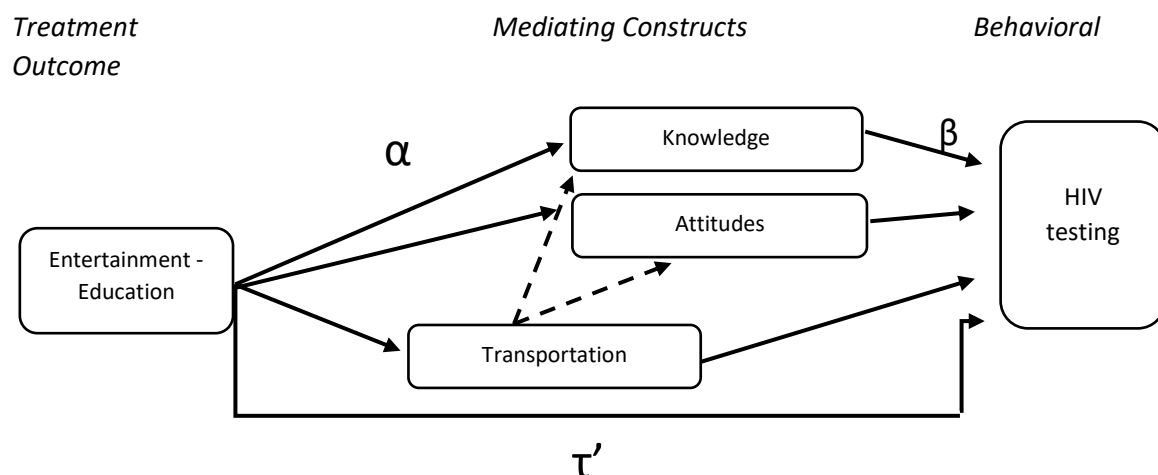
³⁶ For a "short-term" measure collected at the eight-month follow up survey, MTV Shuga increased HIV testing by 2.5 percentage points or 29 percent with respect to the control mean. For a "medium-term" measure collected after the follow up survey (8 to 10 months after the intervention), the television drama increased this behavioral outcome by 3.3 percentage points or 94 percent increase with respect to the control mean.

³⁷ The indirect effect of a mediator on an outcome is commonly calculated as the product ($\alpha\beta$) of the path between the intervention and the hypothesized mediator (α) and the path between the hypothesized mediator and the outcome (β). Direct effects (τ') measure the influence of other mediators not captured by the tested mediators. The sum of direct and indirect effects equals the total effects on the outcome. The direction of the direct effects can be the same or the opposite of indirect effects.

³⁸ This reasonable assumption can be challenged from a stricter point of view for the short-term measure. Impacts on this and the three mediator variables were collected in the same data collection round, the eight-month follow up survey. HIV testing, with its pre- and post-test counseling, is likely to increase HIV knowledge and potentially improve attitudes for individuals that got tested prior to the follow-up survey. Because this additional effect on mediators can endogenously inflate mediation quantities, indirect effects for both psychosocial constructs should be read as upper-bound estimates in the short-term model. This potential inflation issue may also affect the transportation variable, though probably to a lesser extent.

Given the objective of studying the three putative mediators within a causal mediation framework, multi-stage relationships between intermediate variables were not considered (e.g., whether increased HIV testing was mediated by gains in knowledge or attitudes which in turn were mediated by greater transportation). Multi-stage mediation is hard to interpret in a causal framework as its pathways are not always conceptually clear (e.g., improved knowledge could have directly increased HIV testing or indirectly done so through improved attitudes). Behavioral intentions were excluded to focus on realized behavior change. The simplifying assumption of no multi-stage relationships between mediators is likely to underestimate the relative influence of transportation³⁹.

Figure 6. Testing high-level constructs of the theory of change of entertainment-education



To estimate the indirect causal effect for each hypothesized mediator, three single mediator models were conducted (Hicks and Tingly 2011). To estimate (non-causal)

Transportation items were unrelated to HIV testing and the timing of most transportation items focused while watching the television drama.

³⁹ Individuals that report being more emotionally engaged with the television drama and characters are theoretically more likely to absorb embedded HIV messages and to change their attitudes (Hall et al. 2011; Moyer Gusé 2008). This hypothesis is corroborated with the data, as HIV knowledge and attitudes outcomes were statistically mediated by greater transportation by 14 and 26 percent, respectively (*dotted lines* in Figure 2). Large indirect effects are rare in psychology because of the prevalence of multiple mediators (Preacher and Kelly 2011).

indirect effects, one multiple mediator model (Preacher and Hayes 2008) was conducted. The latter takes into consideration the mediating influence of the other two tested variables. It also helps assess how sensitive study estimates of indirect effects are to different methodologies (Lu and White 2014).

Single Mediator models

Mediation analysis in the context of impact evaluation research measures how much of the treatment variable is transmitted by the mediating variable (i.e., impacts on psychosocial and entertainment-education) to the final outcome (i.e., impacts on HIV testing). The below equations present the mediation framework for the case of an RCT. Although the estimation of treatment impacts on Y_i outcomes (equation 1) and of treatment impacts on M_i mediators (equation 2) pose no confounding issues given the random assignment of the treatment (T_i); the estimation of the mediation effects on final outcomes (equation 3) usually violates the unconfoundedness assumption as e_{i2} and e_{i3} are likely to be correlated (Rosenbaum 2002). When the unconfoundedness assumption does not hold, causal estimates will be biased due to omitted variable bias.

1. $Y_i = T_i + e_{i1}$
2. $M_i = T_i + e_{i3}$
3. $Y_i = T_i + M_i + e_{i2}$

The mediation paper uses the approach developed by Hicks and Tingly (2011). This approach was derived using the potential outcomes framework (Holland 1986), commonly used in causal analysis. This method aims to identify the average causal mediation effect (ACME) or causal indirect effect of potential mediators. The ACME is measured as the change in the outcome corresponding to a change in the mediator from

the value that would be realized under the control condition, to the value that would be observed under the treatment condition, while holding the treatment status constant⁴⁰. However, the Hicks and Tingley (2011) framework still does not produce a counterfactual outcome for the mediator as we can't observe potential outcomes for controls. The lack of counterfactuals for these indirect effects prevent experimental designs from identifying the causal impacts of mediators. Thus, identifying causal mechanisms require the sequential ignorability assumption to hold (Rosenbaum 2002). This assumption applies the unconfoundedness assumption for the treatment assignment followed by the observed mediator, given the treatment status and *pretreatment* confounders. The violation to the unconfoundedness assumption is problematic when mediation estimates are given a causal interpretation. This is usually not the case for most mediation studies, as they are based on observational and not experimental data⁴¹. However, this causal mediation approach assumes that the causal relationship between each mediator variable and the outcome is not confounded by the other mediators post-treatment (e.g.,

⁴⁰ In an experimental study, units i are randomly allocated to a treatment ($T_i=1$) or a control condition ($T_i=0$). Because individuals cannot be in both groups, treatment effects ($Y_t - Y_c$) are only observed over a population. Thus, the average treatment effect over a population is defined as $E\{Y_i(1) - Y_i(0)\}$. Mediation analysis aims to quantify the effect of a treatment that operates through a mediator. Let $M_i(t)$ denote the potential value of a given mediator for unit i under the treatment status $T_i=t$; and $Y_i(t, m)$ denote the potential outcome if the treatment and mediating variables equal to t and m . ACME aims to measure how much of the treatment variable is transmitted by the mediating variable. ACME is defined as $d_i(t) = Y_i\{t, M_i(1)\} - Y_i\{t, M_i(0)\}$ for each treatment status $t=0,1$. The causal quantity is the change in the outcome corresponding to a change in the mediator from the value that would be realized under the control condition, $M_i(0)$, to the value that would be observed under the treatment condition, $M_i(1)$, while holding the treatment status constant at t . The average direct effect (ADE) is defined as $c_i = Y_i\{1, M_i(t)\} - Y_i\{0, M_i(t)\}$ for each unit i and each treatment status $t=0,1$. As with treatment effects, we are interested in the average causal mediator effect (ACME) and ADE. Average effects are defined the following way: $ACME = E\{Y_i\{t, M_i(1)\} - Y_i\{t, M_i(0)\}\}$ and $ADE = E\{Y_i\{1, M_i(t)\} - Y_i\{0, M_i(t)\}\}$. However, this framework still doesn't produce a counterfactual outcome for the mediator. While we observe $Y_i\{t, M_i(t)\}$, we don't observe $Y_i\{t, M_i(1-t)\}$ for research designs with one observation per unit. The lack of counterfactuals for ACME and ADE prevents experimental designs from identifying their impacts. Thus, identifying causal mechanisms require the sequential ignorability assumption.

⁴¹ Another important limitation of commonly used mediation methods is that they do not correctly estimate indirect effects when outcomes are binary (Imai, Keele, and Tingley 2010; Pearl 2011; Kohler, Karlson and Holm 2011).

improved knowledge did not affect attitudes). Thus, the sequential ignorability assumption is hard to be met for high-level constructs, and relies on the causal ordering being correct. The latter cannot be verified in the data and can only be supported by theory (Imai et al. 2011).

The mediation approach of Hicks and Tingly (2011) includes a sensitivity test to determine how robust results are to the violations of the sequential ignorability assumption. This sensitivity test shows what correlation between e_{i2} and e_{i3} is needed for the point estimates of the indirect effects to be undistinguishable from zero. Despite the use of experimental data and of even favorable correlation levels between e_{i2} and e_{i3} under the sensitivity test (i.e., high correlations), it should be noted that mediation analysis greatly remains a correlational exercise. Thus, its results should be interpreted as hypothesis-generating research (Gerber and Green 2012; Bullock, Green and Ha 2010). Despite these limitations, the study of causal mechanisms or mediators is too important from both a research and policy perspective that it is imperative to continue developing methodological tools than abandoning their study (Imai et al. 2011).

To estimate the average causal mediation effect (Hicks and Tingly 2011), three single mediator models were conducted, one for each potential mediating variable. The mediation analysis used probit models for the outcome regressions. Calculation of uncertainty estimates is based on the quasi-Bayesian Monte Carlo approximation (King, Tomz and Wittenberg 2000). Other recent mediation approaches that aim to retrieve causal mediation quantities were considered; however, these are not deemed appropriate for the analysis of binary outcomes (Buis 2011) or lacked sensitivity tests (Buis 2011; Daniel et al. 2011) of the sequential ignorability assumption.

Multiple Mediators model

The multiple mediator analysis was based on structural equations commonly used in mediation studies (Preacher and Hayes 2008); with the indirect calculated (MacKinnon et al. 2002) as the product ($\alpha\beta$) of the path between the intervention and the hypothesized mediator (α) and the path between the hypothesized mediator and the outcome (β). By running the three mediator variables under one model, mediation quantities should be lower with respect to single mediator models provided mediator coefficients run in the same direction. It should be reemphasized that structural equation models (Preacher and Hayes 2008) are built to derive associational relationships; so their point estimates don't necessarily have to coincide with those derived from the average causal mediation approach (Hicks and Tingly 2011).

To estimate mediation effects under the structural equation model (Preacher and Hayes 2008), one multiple mediator model was conducted. Following MacKinnon (et al. 2002), HIV testing was considered to be mediated by the hypothesized mediator if (i) the intervention had a statistically effect on the hypothesized mediator (α), the hypothesized mediator was associated with the behavioral outcome (β) after controlling for the intervention effect, and (iii) the mediated effect was statistically significant ($\alpha\beta$). All point estimates report 95% confidence intervals and were based on two-sided statistical tests.

CHAPTER 3. PAPER 1

The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017

Victor Orozco-Olvera^{1*}, Fuyuan Shen², Lucie Cluver³

Keywords: HIV prevention intervention; behavioral interventions; sexual behaviors; mass media; adolescents and youth.

1 Development Research Group. The World Bank. Washington DC, USA

2 Bellisario College of Communications, The Pennsylvania State University, State College, Pennsylvania, USA

3 Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

*Corresponding author

Abstract

Background: Risky sexual behaviors are associated with the transmission of sexually transmitted infections (STIs) and unwanted pregnancies, both major health concerns for youth worldwide. This review studies the effectiveness of narrated mass media programs in promoting safer sexual practices among youth in developed and developing countries.

Methods: Electronic and manual searches were conducted to identify experimental and quasi-experimental studies with robust counterfactual designs published between 1985 and the first quarter of 2017. Effect sizes were meta-analyzed using random-effects models.

Results: Eight experimental and two quasi-experimental studies met our inclusion criteria. The aggregated sample size was 23,476 participants, with a median of 902 participants per study. Entertainment-education narratives had small but significant effects for three sexual behaviors. It reduced the number of sexual partners [standardized mean difference, (SMD)=0.17, 95% confidence interval (CI)=0.02-0.33, three effect sizes], reduced unprotected sex (SMD=0.08, 95% CI=0.03-0.12, nine effect sizes), and increased testing and management for STIs (SMD=0.29, 95% CI=0.11-0.46, two effect sizes). The interventions were not effective in reducing inter-generational sex, measured through the age-gap with sexual partners (SMD=0.06, 95% CI=-0.06-0.19, four effect sizes). Entertainment-education had medium-size effects on knowledge outcomes (SMD=0.67, 95% CI=0.32-1.02, seven effect sizes), where a time-decay relationship is observed. No effects were found on attitudes.

Conclusion: Although mass media entertainment had small effects in promoting safer sexual practices, its economies of scales over face-to-face interventions suggest its potential to be a cost-effective tool above an audience threshold. The use of study

participants from the general youth population and the use of mostly effectiveness trials mitigate concerns regarding its scalability. The overall paucity of high-quality studies affirms the need for strengthening the evidence base of entertainment-education. Future research should be undertaken to understand the moderator effects for different subgroups and intervention characteristics.

Introduction

Risky sexual behaviors are associated with the transmission of STIs and unwanted pregnancies, both major health concerns for youth worldwide (UNFPA 2013). Sexual intercourse remains the main transmission mechanism of HIV among youth (UNAIDS 2016; UNICEF 2016), with inter-generational sex between young women and older men being an important driver in low-income contexts (Dupas 2011; Cluver et al. 2016). Teenage pregnancies are associated with reduction in socio-economic wellbeing of the mother and her children (Cluver et al. 2016; Baird et al. 2014; Toska et al. 2015; Baird et al. 2012; Pennman-Aguilar et al. 2013). Systematic reviews of HIV prevention interventions suggest that while these are effective in developed countries, they are usually not in developing countries (Fonner et al. 2014; Harrison et al. 2010; Krisnaratne et al. 2016; Mavedzenge et al. 2011; Michielsen et al. 2010; McCoy et al. 2010). With HIV prevention funds declining in the last decade (Lei et al. 2015), it is important to invest in effective prevention programs.

Government and development partners have used mass media to persuade youth to engage in safer sexual practices (Brown and Keller 2000; Singhal and Rogers 1999). Mass media entertainment in a narrative format (e.g., dramas, sitcoms), in contrast to informational programs delivered by media (e.g., television public service announcements) is a type of entertainment-education. Though the latter is traditionally defined as the purposeful use of mass media to promote development objectives (Singhal and Rogers 1999; Bandura 2004), entertainment-education also includes commercial productions with prosocial messages that lack social objectives. Social learning theory (Bandura 1977), the main psychological theory underpinning entertainment-education

narratives, argues that show characters may be potential role models; and that they may help improve audiences' self-efficacy for adopting new behaviors (Bandura 2002). Other theories emphasize its potential to reduce counter-arguing when content may be counter-attitudinal (Hall and Bracken 2011; Sood 2002) and updating perceptions of prevalent social norms (Singhal et al. 2004; La Ferrara 2016). Entertainment-education has received increased attention in international development given the limited effectiveness of traditional behavior change programs across different fields, including HIV (World Bank 2015).

It is a longstanding debate on how exposure to entertainment media may affect sexual behaviors (Lowry and Shidler 1993; Page et al. 2000). While observational evidence suggests a relation (Haferkamp 1999; Walsch-Childers and Brown 1993; Anita el 2008), determining whether such exposure influences beliefs and behaviors with observational data is empirically difficult due to sample-selection issues (Heckman 1979; Kearney and Levine 2015). The sexual beliefs and behaviors of a program's audience may differ from those of individuals who decide not to watch the show; preventing studies to measure credible impact estimates. In the last decade, a series of evaluations have emerged that better address selection issues (DellaVigna and La Ferrara 2015; Paluck and Green 2009; Berg and Zia 2015). Although sexual health is among the most common applications of entertainment-education (Pappas-DeLuca et al. 2008; Singhal 2001), the evidence base remains mostly qualitative (Moyer-Guse and Nabi 2011). Previous systematic reviews concluded that mass media interventions had small to medium effects on HIV testing (Vidanapathirana et al. 2005) and risky sexual behaviors (Bertrand et al. 2006; Lacroix et

al. 2014; Shen and Han 2014). These reviews highlight the lack of high-quality evidence in this field.

This meta-analysis aims to assess the effectiveness of entertainment-education narratives on promoting HIV testing and reducing risky sexual behaviors of youth in developed and developing countries. Youth was defined as individuals between 15 and 24 years of age. The scope of this review is narrower with respect to previous reviews of mass media interventions with sexual health objectives. This review will (i) be restricted to the mentioned age group, as its psychological development, relationship characteristics and sexual behaviors is likely to differ from those of older age cohorts (e.g., single individuals are likely to have a greater number of sexual partners, where condom use is more relevant); (ii) include experimental and quasi-experimental studies that have strong identification strategies following Cochrane Review guidelines (i.e., we exclude studies where selection bias is likely to persist); (iii) interventions that have a narrative format (e.g., we exclude for example flyers and public service announcements that lack this format); (iv) preferably include effectiveness trials where the type of program exposure follows that of the scaled-up intervention. For example, “voluntary” take-up for community screenings or “imposed” exposure for school-based screenings; and (v) studies that measured actual rather than intended behavior change by excluding studies that collected post-intervention data immediately after program exposure.

Methods

This review follows reporting standards for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al. 2009). The PRISMA checklist is in S1

Table. This review was registered in the PROSPERO International prospective register of systematic reviews (PROSPERO 2016: CRD42016046005).

Selection criteria

Following Cochrane Review standards, studies were eligible for inclusion if they were randomized controlled trials (RCT), cluster RCTs, and quasi-experimental studies with robust counterfactual designs, including interrupted time series designs with control groups. We included studies that measured outcomes at least one week after exposure. Studies of complex interventions (those that included other information components, such as HIV information) were included if the edutainment component played a prominent part. By homogenizing the interventions targeting youth, the review aims to better identify the effectiveness of the narrated or storytelling format.

Search Strategy

Published studies were searched in English. The search period covered 1985 to the latest available year. We completed our searches on March 31, 2017. The following databases were searched: Communication & Mass Media Complete, EBSCO, ERIC (CSA), EconLit, JSTOR, MEDLINE, CINAHL (EBSCO), JSTOR, metaRegister of controlled trials, NBER, OpenSIGLE (grey literature), PsycINFO, RePEC, and The Cochrane libraries. The full search strategy of the MEDLINE database is provided as S1 List. References from previous reviews and from retrieved articles were examined for additional studies. Journal articles, published reports and for the case of economics research, published working papers were included (prior to journal publication, the latter are standard practice in the economics literature). Authors were contacted for further information when required.

Risk of bias assessment

Study quality was assessed using the Cochrane Review guidelines for assessing risk of bias (Higgins and Greene 2011). Each study was evaluated for participant selection (sequence generation and allocation concealment), blinding (personnel and outcome assessors), attrition (> 80% retention), and selective reporting and other types of biases. Each item was scored as either low, high or unclear risk of bias.

Data analysis

For cluster RCTs or natural experiments delivered that failed to cluster their standard errors; we used a conservative value for the intra-class correlation coefficients of 0.05 (Glassman et al. 2014). We preferred intent-to-treat rather than per protocol estimates to minimize potential selection issues.

Standard meta-analytical methods were used. We used the Comprehensive Meta-Analysis software to convert all binary and continuous data outcomes to standardized mean differences and their 95% confidence intervals (DerSimonian and Laird 1986; Wilson and Lipsey 2001). The data was coded so that safer sexual behaviors translated into positive values. For studies that reported different types of primary behaviors, outcome values were aggregated into four behavior categories: (i) age-gap with sexual partners; (ii) number of sexual partners (iii) unprotected sex; and (iv) STI testing and follow up.

The primary analysis for each of the four behaviors includes some degree of heterogeneity as it synthesizes data from non-identical interventions and from similar but different outcome indicators. Study heterogeneity is recognized by applying the random-effects

model (Hedges and Vevea 1998) and by calculating the I² statistics (The Cochrane Collaboration 2011; Higgins et al. 2003). Our random-effects models report statistics based on random-effects weights within and not between subgroups. The meta-analysis used the study as the unit of analysis and the mean of the selected outcomes (we did not assume independence of outcomes within each study). Each included study only had one-time point. The I² statistic describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error.

To determine whether results were robust to methodological decisions, we re-ran the analysis for the primary outcomes by repeating the meta-analysis using fixed-effects models; we removed one study at a time to determine if it substantially affected the meta-analyzed effect size, and we re-ran the analysis without studies that were rated with high risk-of bias (defined as having more than two out of seven items evaluated at high-risk). To further assess the likely non-independence of outcomes within a study, we ran a random-effects model combining all behavior outcomes. We assessed whether publication bias was present through funnel plots (Egger et al. 1997) and Rosenthal's (Rosenthal 1979) and Egger's tests (Sterne et al. 2001).

Results

Results of the search

As seen in Figure 1, our electronic search strategy yielded 2,450 records. Thirty-three additional records were identified through the review of reference sections of relevant studies and of previous systematic reviews. Upon review of study titles and abstracts, 2,340 studies were excluded. The leading exclusion reasons were (i) the study did not include both entertainment-education and sexual health objectives (e.g., the study excluded one of them) or (ii) it was theoretical or qualitative in nature, so the study did not aim to measure program impacts on sexual health. Forty-six studies were deemed as quantitative evaluations that suggested causal analysis. Of these, the complete article was retrieved and assessed. Thirty-six studies did not meet our inclusion criteria. S2 List and S2 Table respectively provide the list and the reasons why they these studies were excluded. Thus, the meta-analysis is based on 10 studies. Eight studies included behavior outcomes and two studies only secondary outcomes (knowledge and attitudes). Study overlap with previous systematic reviews of mass media interventions was low: two of the ten included studies (Mileliri et al. 2003; Vaughan et al. 2000) were covered in at least one of the previous reviews.

Key features of included studies

Table 1 presents information regarding the studied populations and the interventions of the 10 studies. S3 Table provides information regarding their study designs and their outcomes. Included studies were reported between 1988 and 2017, with half of them reported 2011 onwards. Five studies took place in the US and five in Sub-Saharan Africa. The country's HIV prevalence rate when the studies took place varied between 0.6 and

7.0 percent. The unweighted average age across studies was 19.3 years. Most included studies took place in urban settings. The studied interventions included TV screenings (n=5), TV broadcasts (n=1), radio broadcasts (n=1), videos delivered through smart phones or personal computers (n=2), and comics (n=1).

While the experimental studies had control over program exposure, the non-experimental (field) evaluations were based on potential exposure to public broadcasts. Controlled exposure was usually under one hour, while potential exposure was above 40 hours. As effectiveness trials, the meta-sample was large with a total of 23,476 participants (a median of 902 participants per study). The endline outcomes were measured between two weeks and one year, with six studies measuring outcomes after three months or more.

Figure 1. Flow Diagram of study selection process, 1985-2017

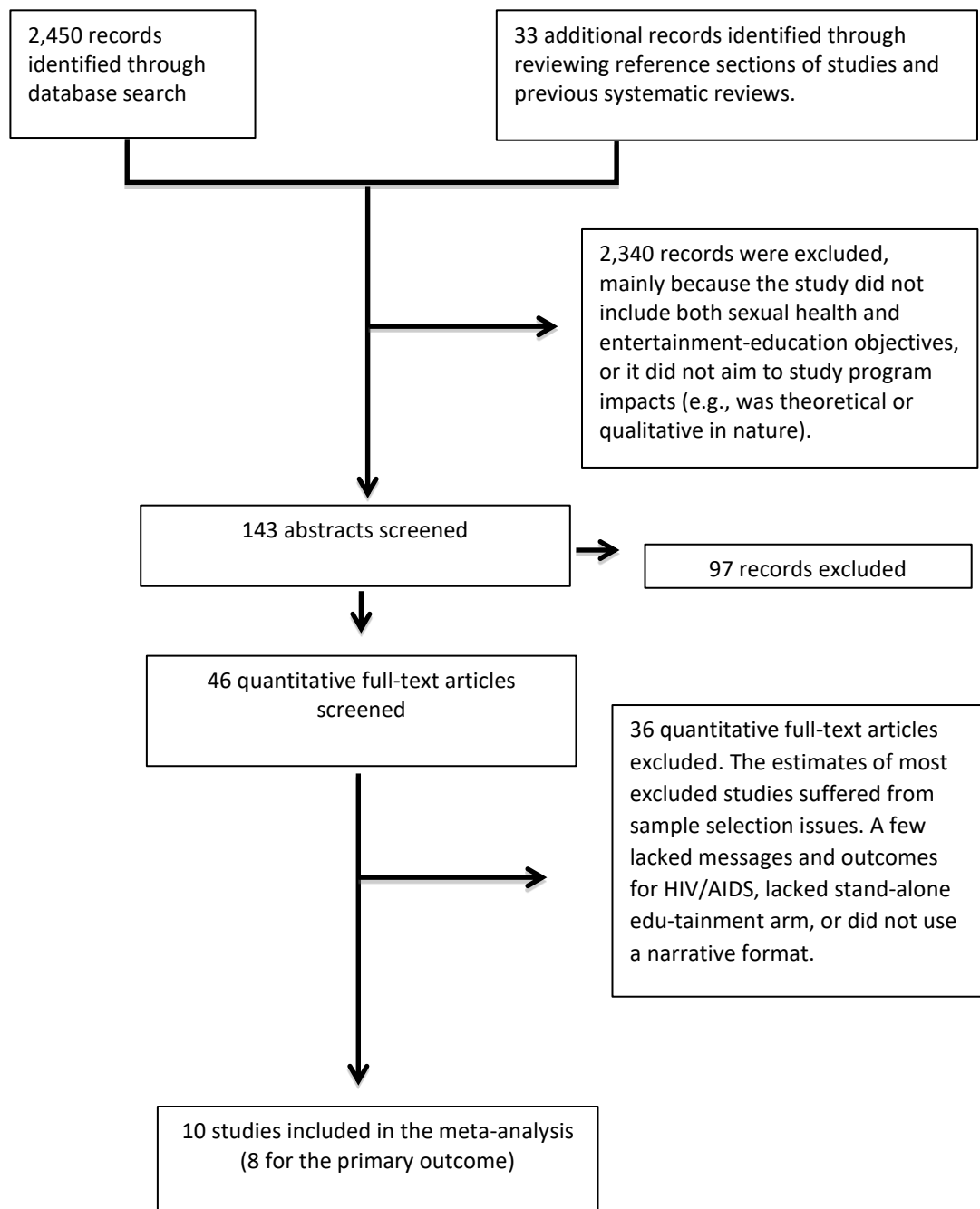


Table 1. Descriptive Characteristics of 10 included studies

Study	Study population Characteristics	Sexual behavior included	Sample size	Intervention and Exposure time	Follow up
Banerjee <i>et al.</i> 2017	Men and women (average age 20.5) in urban Nigeria	Yes	4,986	Community screenings of edu-tainment TV drama season, MTV Shuga (2.5 hours)	6 months
Dupas <i>et al.</i> 2011	Eighth-grade school female students (average age 15.3) in rural Kenya	Yes	6,196	UNICEF edu-tainment comic movie “Sara: The Trap” (10 minutes), combined with HIV relative risk information	9-12 months
Dupas <i>et al.</i> 2017	Eighth-grade school female students (average age 15.5) in rural and urban Cameroon	Yes	700	Same than Dupas 2011. This complex intervention was experimentally added to a broader intervention	9-12 months
Jones <i>et al.</i> 2012	African American women (average age 22) in urban US	Yes	902	Twelve 18-minute episodes of edu-tainment TV drama “Love, Sex, and Choices”, streamed through smart phones (3.5 hours)	3 months
Kearney <i>et al.</i> 2015	Female teenagers (average age 17) in rural and urban US	Yes	4,100*	TV broadcasts of reality commercial series “MTV 16 and Pregnant” (potential exposure of 45 hours)	3 months
Milleliri <i>et al.</i> 2003	Male and female students (average age 19.1) in peri-urban Gabon	No	594	Cartoon comic about risks of unprotected.	3 weeks
Moyer-Gusé <i>et al.</i> 2010	Male and female college students (average age 20) in urban US	Yes	360	Episode screening of TV commercial drama “The OC” (30 mins)	2 weeks
Solomon <i>et al.</i> 1988	African American male patients of STI clinic (average age 23) in urban US	Yes	902	Educational video clip in a drama-format on top of counselling about importance of STI treatment (10 mins)	2 weeks
Vaughan <i>et al.</i> 2000	Men and women (average age 31) in rural Tanzania	Yes	1,230	Bi-weekly radio broadcasts of edu-tainment show “Let's Go with the Times” for two years (potential exposure of 50 hours)	3 months
Wang <i>et al.</i> 2016	Female undergraduate Latina students (average age 20.7) in urban US	No	62	Online exposure to TV show East Los High (potential exposure of one hour).	2 weeks

*Number of participants is larger as the number represents instead development media markets (205, across 20 quarters for a total of 4,100 observations).

Regarding the study design and quality, eight studies were experimental and two used quasi-experimental methods. The support for risk-of-bias judgements is presented in S3 Table. The most common risk-of-bias was not fully blinding participants and assessors, followed by incomplete outcome data at follow-up. The overall risk of bias is considered low for most studies.

The data suggests that publication bias is likely to be present, though it should not substantially change the estimates of our effect sizes. On the one hand, the funnel plot shows a higher concentration of studies with positive results (See S1 Fig). On the other hand, Egger's Test of the Intercept [Egger's Intercept (2 tails)=1.54, 95% CI=-0.34-3.44, $t=1.99$, $df=6$, $p=0.10$] and Rosenthal's fail-safe ($N=45$) suggest no publication bias. The latter test means that we would need to locate and include 45 null studies (or 5.6 studies null studies per included study) for the combined 2-tailed p-value to exceed 0.05. Despite the large meta-sample ($n=23,476$), the results of this review should be read cautiously given the small number of studies in the primary analysis ($n=8$).

Measures of risky sexual behavior

Five studies collected objective behavior measures, usually in addition to the self-reported data. Objective measures included teenage birth rates, information-seeking behaviors, and percent of patients that got tested for HIV. Most self-reported items were similar to HIV questions used by the Demographic and Health Surveys program (ICF International 2016), a standardized health survey conducted in over one hundred countries. Fig 2 presents the effect sizes for the four behavior categories, and the overall pooled effect. S4 Table lists the SMD effects for individual behavior outcomes.

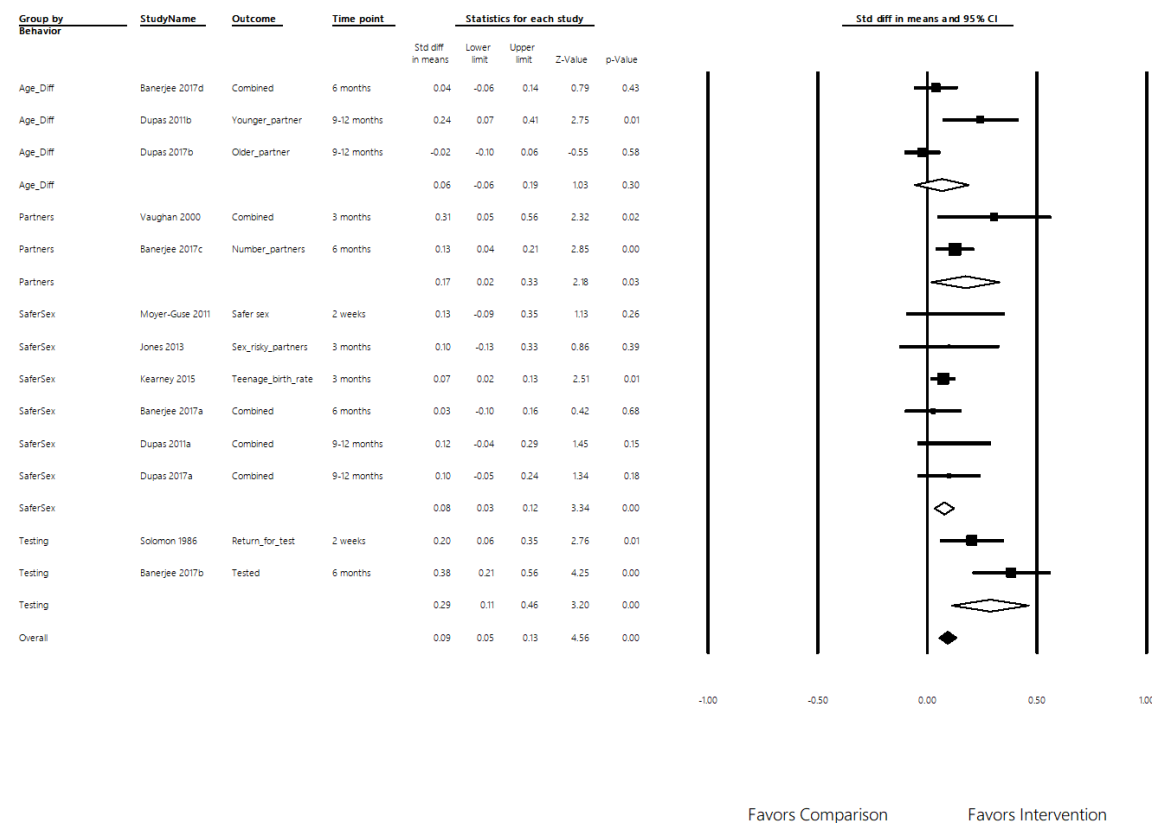
Effectiveness on Safer Sexual Practices

Overall, individuals that were exposed to entertainment-education narratives were more likely than comparison participants to engage in safer sexual practices. We observed clinically small (SMDs between 0.08 and 0.29) but statistically significant effects for three of the four measured sexual behaviors. Three studies measured impacts on inter-generational sex, through the age-gap of sexual partners (Dupas 2011; Banerjee, La Ferrara and Orozco 2018; Dupas, Huillery and Seban 2017). The effect size was not significant (SMD=0.06, 95% CI=-0.06-0.19, four effect sizes), though study heterogeneity was high ($Q=7.5$, $df(Q)=2$, $I^2=73.3$). Two studies measured effects on the number of sexual partners (Vaughan et al. 2000; Banerjee, La Ferrara and Orozco 2018). The effect size was significant (SMD=0.17, 95% CI=0.02-0.33, three effect sizes), though study heterogeneity was high ($Q=1.7$, $df(Q)=1$, $I^2=40.3$). Six studies measured unprotected sex practices (Dupas 2011; Kearney and Levine 2015; Moyer-Guse and Nabi 2011; Banerjee, La Ferrara and Orozco 2018; Dupas, Huillery and Seban 2017; Jones, Lacroix and Hoover 2012), defined by this review as measures related condom use and pregnancy. The effect size for unprotected sex was significant (SMD=0.08, 95% CI=0.03-0.12, nine effect sizes), with no statistical heterogeneity observed ($Q=1.2$, $df(Q)=5$, $I^2=0$). Two studies measured STI testing and management follow up practices (Banerjee, La Ferrara and Orozco 2018; Solomon and DeJong 1988), finding a significant effect (SMD=0.29, 95% CI=0.11-0.46, two effect sizes). High heterogeneity was observed ($Q=2.4$, $df(Q)=1$, $I^2=57.9$).

The pooled effect of the four behavior outcomes was small and significant (SMD=0.09, 95% CI=0.06-0.12, eighteen effect sizes). Though the pooling of these different behaviors may be difficult to conceptually justify, this pooled approach is used in sensitivity tests

and subgroup analysis given the limited number of studies. Sensitivity tests did not reveal single studies exerting influence on the overall effect size or effects changing by removing studies assessed at high risk of bias. The pooled effect did not differ when pooling all behavior outcomes using a random-effects model.

Figure 2. Effects on four sexual behaviors. Age-gap, number of sexual partners, unprotected sex, and STI testing and follow up. Positive values represent safer sexual practices. Standardized Mean Differences of Behavior Outcomes, 95% Confidence Interval.



Meta-analysis Subgroup analysis

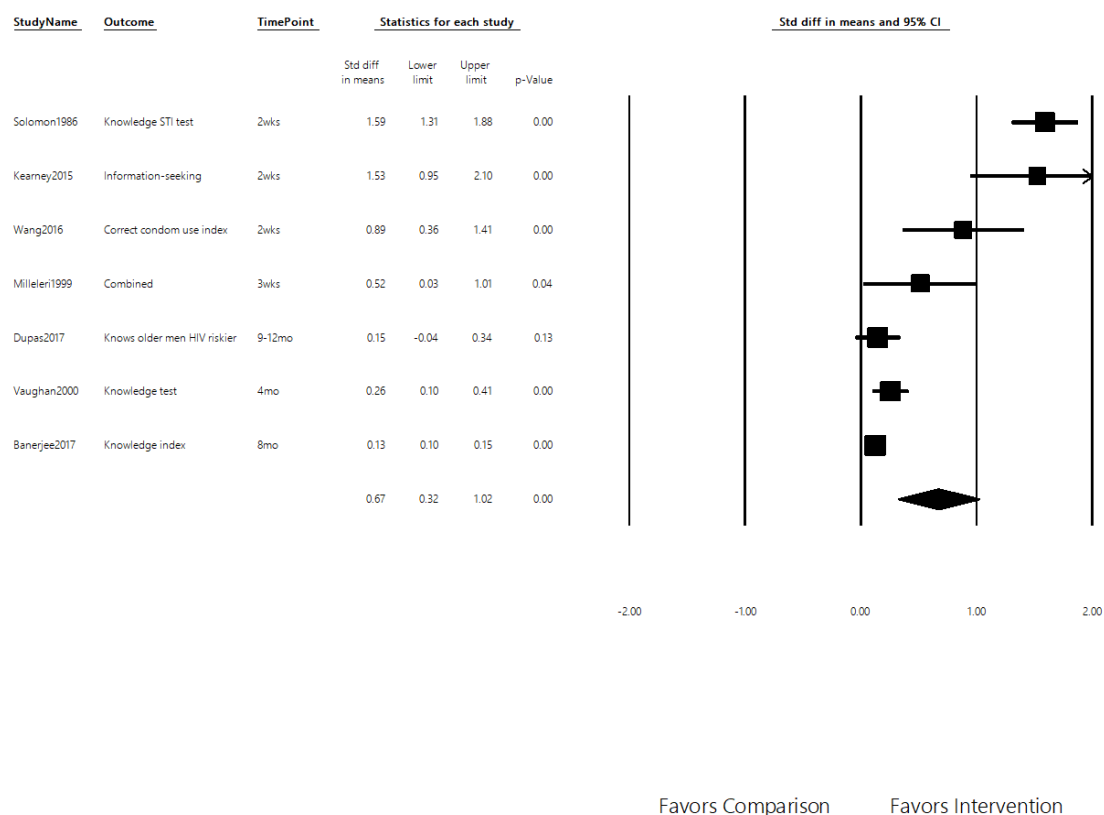
Because factors co-vary across subgroups in meta-analyses, the limited number of included studies prevented conducting a moderators analysis related to study design, key subgroups and intervention characteristics. We present the moderation effects of the timing of the post-intervention survey due to its clearer theoretical underpinnings (i.e., we would expect program effects to diminish with time). The timing of the post-intervention data collection did not moderate outcomes, providing indicative evidence of medium-term behavior change. A meta-regression with respect to the timing of the follow up survey finds a negative and but not significant relation ($Q=1.19$, $df(Q)=1$, $p=0.27$). The graph of this meta-regression can be found in S2 Fig.

Knowledge and Attitudes

Ten knowledge outcomes were retrieved from seven studies. Measurements for most studies were comprehensive, though a few were restricted to measuring narrow outcomes related to program messages. A medium and significant effect was found for knowledge outcomes ($SMD=0.67$, $95\% CI=0.32-1.02$). As seen in Fig 3, the effect size declined with the timing of the follow-up survey.

This time-decaying relationship for knowledge outcomes is confirmed by the subgroup analysis. The effect size for studies that collected post-intervention data after a month ($SMD=0.15$, $95\% CI=0.08-0.22$) was approximately a fifth of the effect size of studies that measured effects within a month ($SMD=1.14$, $95\% CI=0.61-1.68$), with the difference in effect size being statistically significant ($p<0.01$).

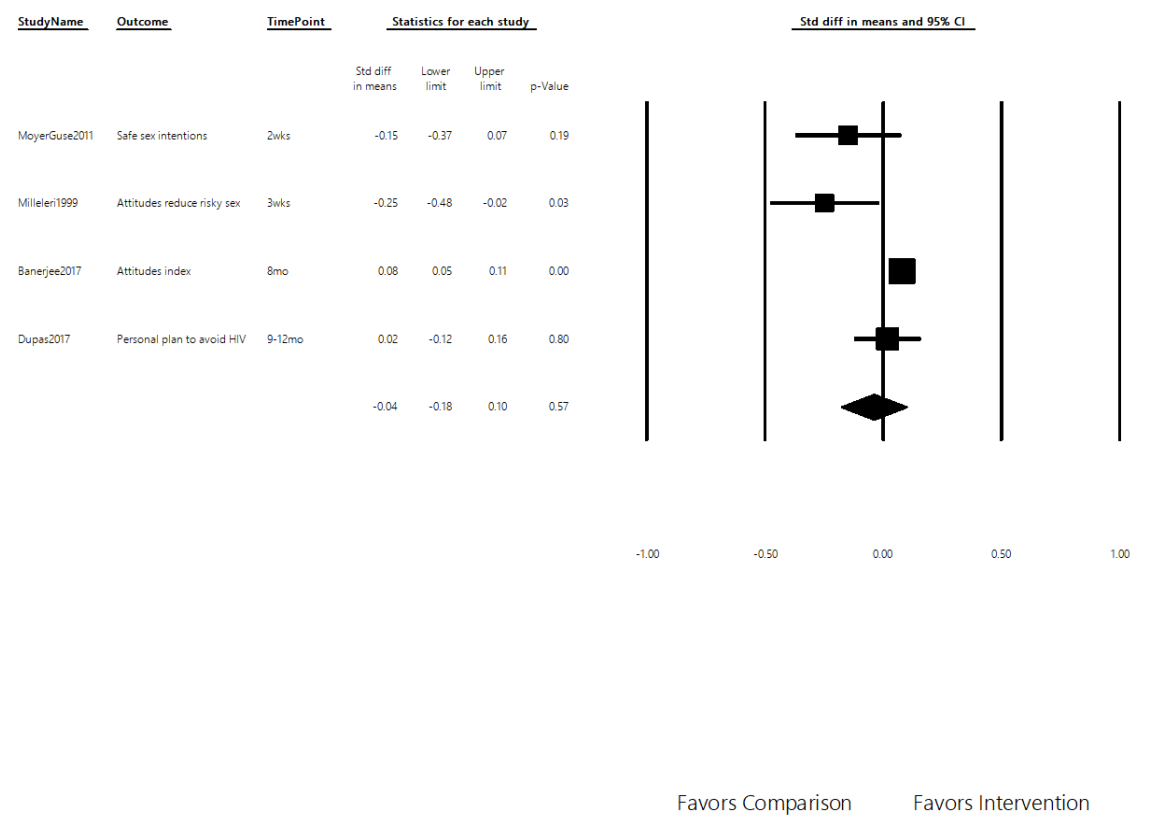
Figure 3. Effects on knowledge. Positive values represent intended effects. Standardized Mean Differences of Secondary Outcomes, 95% Confidence Interval.



Four attitudes outcomes were retrieved from four studies. All included studies measured prevention intentions, with one study (Banerjee, La Ferrara and Orozco 2018) also measuring attitudes towards HIV testing and people living with HIV/AIDS. As seen in Fig 4, while no effects were observed for the pooled estimate, small negative effects were observed for the studies that measured outcomes within a month (SMD=0.20, 95% CI=-0.36-0.04) and small positive effects when attitudes were measured after a month (SMD=0.08, 95% CI=0.05-0.10). The contradictory finding of the former is consistent with several excluded studies that collected data immediately after program exposure. Those studies found significant negative effects on behavioral intentions and attitudes. This may

suggest self-reported biases driving this result rather than actual negative short-term effects on attitudes (Bensley and Wu 1991).

Figure 4. Effects on attitudes. Positive values represent intended effects. Standardized Mean Differences of Secondary Outcomes, 95% Confidence Interval.



Quality of the Evidence

Using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria (GRADE Working Group 2004), the quality of the evidence was considered moderate for inter-generational sex and unprotected sex; and low for number of sexual partners and STI testing and management. A “moderate” quality rating suggests further research is likely (very likely for a “low” rating) to have an important impact on our confidence in the effect size estimate. For knowledge and attitudes, it was rated high and very low respectively. See S5 Table for the GRADE Summary of Findings table. These

ratings may be conservative, as recent research (Movsisyan, Melendez-Torres and Montgomery 2016) shows that GRADE guidelines rarely give high ratings to complex and field interventions.

Discussion

This review found that entertainment-education narratives were effective in promoting HIV testing and reducing risky sex among 15-24-year-old youth. Results were robust to a series of methodological decisions. Behavioral effects persisted for several months after program exposure, with no distinguishable differences between short and medium-term effects. Behavioral effects were small (SMD between 0.08 and 0.29) and were approximately half the size reported in previous meta-analyses of mass media interventions (Lacroix et al. 2014; Shen and Han 2014). However, a direct comparison is not possible given important differences in terms of populations, outcomes, study designs and interventions. The use of study participants from the general population and the use of effectiveness trials mitigate concerns regarding the scalability of entertainment-education narratives (Singhal and Rogers 2002). Not clearly reporting blinding and selective reporting of outcomes, were the main quality concerns of included studies. Despite their relatively large sample sizes and high-quality designs, the limited number of studies remains the main limitation of this meta-analysis. Another study limitation was the inconsistency and the imprecision of effects across studies, which reduced the GRADE quality of evidence score despite the overall high quality of included studies.

Systematic reviews of HIV behavior change find random effects on reducing risky sex and HIV infections, especially in developing countries (Krishnaratne et al. 2016; Fonner et al. 2014). Despite the evidence of small size effects, the marginal cost for reaching individuals

should be lower above a certain “audience” threshold when compared to more resource-intensive or face-to-face interventions. High-quality entertainment shows are often consumed by millions of individuals. For example, one of the studied interventions (Banerjee, La Ferrara and Orozco 2018) is broadcast in over seventy countries. Most included studies that found positive effects in knowledge outcomes also found behavior effects, which suggests that knowledge is a necessary but not a sufficient condition for behavior change. The ambiguous impacts on attitudes may be partly explained by measurement issues, including reporting biases triggered by short-term surveys that make salient program objectives (Bensley and Wu 1991). Data limitations prevented carrying out a subgroup analysis for key subgroups and intervention characteristics.

Although the number of higher-quality evaluations has been slowly growing in the last decade, important knowledge gaps remain. More experimental research is needed, particularly in Europe, Latin America and South Asia. This review could not find studies in these regions that met our inclusion criteria. The question of program effectiveness per HIV prevalence or country-income levels remains. Moderators research is needed to better understand the dose-response relationship and the minimum amount of investments needed to expect effects in the medium-term. Studies should be powered to study impacts for different subgroups (Cho et al. 2013; Baird et al. 2014), including gender, urban/rural location and audiences’ exposure to media. Of particular policy interest is understanding program effects on individuals that repeatedly engage in risky sexual behaviors as these individuals have the potential to infect a larger number of individuals (McCoy et al. 2010). Researchers may also experimentally investigate the effectiveness of different media outlets (i.e., TV, radio, smartphones); and study the additional

effectiveness of individual components of comprehensive media campaigns (Coville, Orozco and Reichert 2019). Given the presence of social desirability bias in HIV research, future research should complement survey work with objective measurements, such as STI biomarkers.

With the rapid expansion of mass media in developing countries, entertainment-education offers an important opportunity to promote positive behavior (Moreau 2011; World Bank 2016). As development partners are increasingly demanding higher quality evaluations (Banerjee and Duflo 2009; Legovini, Di Maro and Piza 2015), and with competing HIV prevention approaches producing experimental evidence of their investments (de Walque, Dow, Nathan 2012; Baird, McIntosh and Ozler 2011), entertainment-education needs to invest in stronger evaluation research. The evidence base remains mostly qualitative or correlational. Credible program effects may help future investments towards this promising communications approach.

ACKNOWLEDGEMENTS

Financial support: No external sources supported this study. The findings and conclusions are those of the authors only.

Dual Publications: This meta-analysis is based on published or public results of other studies.

Conflicts of interest. The authors have declared that no competing interests exist.

Differences between protocol and review: The review added complex interventions that included an edu-tainment module only when it played a prominent role in the intervention (the protocol had noted their overall exclusion).

Supporting Information-Meta-analysis

S1 Table. PRISMA Checklist (reported page number corresponds to the published paper)

Section/ topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	p. 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	p. 1-5
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	p. 3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	p. 5-7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	p. 5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Table 1, S3 Table
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	p. 6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	S1 List
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Fig 1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	p. 5-6, and 22
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	p. 7, 22, S3 Table

Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	p. 6, 11-12, S3 Table
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	p. 7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	p. 7-8

Section/ topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	p. 11-12
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	p. 15-17
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig 1, S2 List and S2 Table
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 1, S3 Table
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	S3 Table
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Fig 2, S4 Table
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	p. 12-13, Fig 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	S1 Fig
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	p. 7, 13, S2 Fig
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	p. 18-19, S5 Table

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	p. 18
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	p. 19-20
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	p. 22

S2 Table. Intervention and Reason for Exclusion

Study	Intervention	Reason for exclusion
Anita 2008	Effects of TV on teenage pregnancies	The study compares viewers vs. non-viewers without properly addressing selection issues.
Arroyave 2008	Effects of edutainment dramas on sexual attitudes and intentions.	RCT. Based on exit surveys collected immediately after exposure.
Brodie 2001	TV broadcasts of “ER” in the USA with HIV messages	The study sample included a wide range of ages, with the youngest cohort set at 18-30 years of age.
Boulay 2002	Radio drama in Nepal with family planning focus.	The study compares listeners vs. non-listeners with only post-intervention data.
Collins 2003	TV sitcom episode in the USA with condom efficacy message	The study compares viewers vs. non-viewers with only post-intervention data.
Do 2006	TV drama in Bangladesh with family planning focus.	The study uses propensity score matching with 14 control variables only using a large post-intervention survey. Selection bias may persist.
Farr 2005	TV program about HIV/AIDS in Ethiopia	The study compares viewers with non-viewers with limited controls. Selection bias may persist.
Geary 2007	Evaluation of HIV mass media campaigns in Nepal, Brazil and Senegal	The campaign included a series of media components. The study collects data before and after data, comparing consumers vs non-consumers.
Jadranin 2015	Movie with HIV/AIDS messages with Serbian soldiers	Before-after comparison that lacks a control group.
Jensen 2009	Community cable access (main medium for soap operas in India) in India	Family planning outcomes (number of children and birth spacing at the individual level) lacking high-risk sexual behaviors and not provided for youth.
Karlyn 2001	Radio drama in Mozambique aimed to prevent STI	The study compares listeners vs. non-listeners with only post-intervention data.
Keating 2006	Evaluation of HIV mass media campaigns in Nigeria	The campaign included a series of media components. The study collects data before and after data, comparing consumers vs non-consumers.
Kuhlmann 2008	Radio drama in Botswana with PMTCT focus	The study compares listeners vs. non-listeners with only post-intervention data.
La Ferrara 2012	Commercial TV dramas in Brazil	The study did not focus on HIV messages and youth.
Lapinski 2008	Short film about HIV stigma in Nigeria	RCT. Based on exit surveys collected immediately after exposure.
Middlestadt 1995	Radio campaign in St. Vincent and the	The study compares listeners vs. non-listeners with only post-intervention data.

	Grenadines with HIV/AIDS messages	
Mohammed 2001	Radio drama in Tanzania with family planning focus	The study compares listeners vs. non-listeners with only post-intervention data.
Moyer 2011b	TV drama in the USA with safe sex messages	RCT. The treatment arm includes post-screening discussions.
O'Donnell 1995	Video-based interventions promoting condom use in STD clinic	RCT. The intervention video-based patient education program and not an entertainment-education drama
O'Leary 2007	TV drama in Botswana with HIV de-stigmatization focus	The study compares viewers vs. non-viewers with only post-intervention data.
Pappas-DeLuca 2008	Radio drama in Botswana with HIV prevention focus.	The paper studies associations between program exposure and outcomes, without attempting to construct a control group.
Peltzer 2003	TV drama in South Africa with HIV and condom messages	The TV drama was bundled with other interventions. The study is based on post-intervention data only.
Peltzer 2004	TV drama in South Africa with HIV and condom messages	The TV drama was bundled with other interventions. The study is based on post-intervention data only.
Ramirez 2014	Short film for US Latino communities with Stigma towards gay and bisexual	Before-after study without a comparison group
Rogers 1999	Radio drama in Tanzania with family planning messages	The evaluation of this intervention is presented by family planning outcomes (Rogers 1999) and high-risk sexual outcomes (Vaughn 1999). The latter is an included study.
Sahka 2013	TV drama in Iran targeted to FSW	Before-after study that bundled several interventions to the educational movie
Shapiro 2006	TV drama in Cote d'Ivoire with HIV focus	The study compares viewers vs. non-viewers with only post-intervention data.
Silvestre 2015	Evaluation of HIV mass media campaigns in Nepal, Brazil and Senegal	The campaign included a series of media components. The study compares consumers vs non-consumers with only post-intervention data.
Smith 2007a	Radio drama in Ethiopia with HIV prevention focus	The study compares listeners vs. non-listeners with only post-intervention data.
Storey 1999	Radio drama in Nepal with family planning messages	Study targets married women and only includes family planning outcomes.
Vaughan 2000b	Radio drama in St Lucia with HIV prevention focus	The study is underpowered and selection bias is likely to be significant. Although study collected pre and posttest data for listeners and non-listeners, limited attempts to control for characteristics that could affect listenership and key outcomes.

Vernon 2007	Radio drama in Colombia with family planning focus	The study is based on a before-after analysis without a comparison group.
Whittier 2005	TV sitcom episode in the USA with STI focus for MSM	The study compares viewers vs. non-viewers with only post-intervention data.
Yoder 2006	Radio drama in Zambia with HIV prevention focus	The study collects data before and after data, comparing listeners to non-listeners. Selection bias may persist.

S3 Table. Characteristics of Included Studies and Support for Risk of Bias Assessment

Banerjee 2017

Methods	CRCT	
Outcomes	Knowledge (about ARVs and of waiting period), Sexual behaviors (used condom in last sexual intercourse, reduction in number of current sexual partners, got tested in last six months).	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Community centers were randomly allocated with computer random number generator.
Allocation concealment (selection bias)	Low risk	Community centers centrally allocated by the research team.
Blinding of participants and personnel (performance bias)	High risk	Community-level randomization with buffer areas minimize performance bias.
Blinding of outcome assessment (detection bias)	Low risk	In addition to self-reported data, the study collects objective measures for HIV testing. The post-intervention survey was conducted by a survey firm independent of the implementing NGO and six months after exposure, which further reduces concerns for detection bias.
Incomplete outcome data (attrition bias)	Low risk	The study's attrition rate is 3 percent, with no systematic differences in overall rates or reasons between the treatment and the control groups.
Selective reporting (reporting bias)	Low risk	Reporting bias is unlikely given that the study protocol pre-specifies primary and secondary outcomes and the study reports results for all these outcomes, in addition to constructing aggregated indexes.
Other bias	Low risk	Baseline balance, standard errors were properly adjusted in the analysis and recruitment for the treatment and control groups was conducted in a similar way.

Dupas 2011

Methods	CRCT	
Outcomes	Sexual behaviors (teenage pregnancy, used condom last sexual intercourse, age difference with baby's father)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Schools were randomly allocated with computer random number generator.
Allocation concealment (selection bias)	Low risk	Schools centrally allocated by the research team.
Blinding of participants and personnel (performance bias)	Low risk	School-level randomization with buffer areas minimize performance bias.
Blinding of outcome assessment (detection bias)	Low risk	The primary outcome, teenage pregnancy, is objectively corroborated by the enumerators.

Incomplete outcome data (attrition bias)	Low risk	Attrition rates are under 2 percent, with no systematic differences between treatment and control groups.
Selective reporting (reporting bias)	Low risk	All primary outcomes are reported.
Other bias	Low risk	No other issues were found.

Dupas 2017

Methods	CRCT	
Outcomes	Sexual behaviors (teenage pregnancy, used condom last sexual intercourse, age difference with baby's father)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Schools were randomly allocated with computer random number generator.
Allocation concealment (selection bias)	Low risk	Schools centrally allocated by the research team.
Blinding of participants and personnel (performance bias)	Low risk	School-level randomization with buffer areas minimize performance bias.
Blinding of outcome assessment (detection bias)	Low risk	The primary outcome, teenage pregnancy, is objectively corroborated by the enumerators.
Incomplete outcome data (attrition bias)	Low risk	Attrition rates are under 2 percent, with no systematic differences between treatment and control groups.
Selective reporting (reporting bias)	Low risk	All primary outcomes are reported.
Other bias	Low risk	No other issues were found.

Jones 2012

Methods	Quasi-experimental (ITS)	
Outcomes	Knowledge (Knew about prevention options after unprotected sex) and sexual behaviors (Vaginal episode equivalent with high risk partners)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Study participants were randomly allocated with computer random number generator.
Allocation concealment (selection bias)	Low risk	Group assignment placed into sealed envelopes.
Blinding of participants and personnel (performance bias)	Unclear risk	Staff was not blinded though the mobile-based intervention did not require participants to interact with each other.
Blinding of outcome assessment (detection bias)	Low risk	Though based on self-reported data, the study uses audio-computer assisted self-interviews to minimize detection bias.
Incomplete outcome data (attrition bias)	High risk	32/149 and 25/146 missing from intervention and control groups respectively.

Selective reporting (reporting bias)	Unclear risk	The study lacked a research protocol though main indicators seem to be reported.
Other bias	Low risk	No other issues were found.

Kearney 2015

Methods	Quasi-experimental (ITS)	
Outcomes	Knowledge-seeking behavior (Google searches for how to get birth control pills) and Sexual behavior (teenage pregnancy).	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The study is not experimental, though it provides a series of statistical tests to validate the used instrumental variable strategy.
Allocation concealment (selection bias)	Low risk	Natural experiment
Blinding of participants and personnel (performance bias)	Low risk	The study relied on district-level data and did not interact with study participants.
Blinding of outcome assessment (detection bias)	Low risk	All outcomes are measured through administrative and independent data.
Incomplete outcome data (attrition bias)	Low risk	Only one of 205 Designated Market Areas was removed from a 20-quarter panel.
Selective reporting (reporting bias)	Low risk	Though study seemed to have lacked a research protocol, the main outcomes are reported.
Other bias	Low risk	No other issues were found.

Milleleri 1999

Methods	CRCT	
Outcomes	Knowledge (testing and transmission mechanisms) and attitudes (ability to convince peers)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear how schools were randomly allocated.
Allocation concealment (selection bias)	Low risk	The research team centrally allocated schools.
Blinding of participants and personnel (performance bias)	Low risk	School-level randomization with buffer areas minimize performance bias.
Blinding of outcome assessment (detection bias)	Low risk	Knowledge questions less likely to suffer from detections bias.
Incomplete outcome data (attrition bias)	High risk	Attrition rates above ten percent.
Selective reporting (reporting bias)	High risk	The study lacked a research protocol, with key outcomes not reported.

Other bias	Unclear risk	The study does not provide evidence regarding pre-intervention balance.
------------	--------------	---

Moyer-Guse 2010

Methods	RCT	
Outcomes	Sexual behavior (use of birth control)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Study participants were randomized.
Allocation concealment (selection bias)	Unclear risk	Anonymity codes were used to match participants. However, students may have known each other making concealment difficult.
Blinding of participants and personnel (performance bias)	High risk	The lab setting, and watching the intervention in group, probably triggered Hawthorne effects.
Blinding of outcome assessment (detection bias)	High risk	Self-reported data that could be linked with the intervention objective.
Incomplete outcome data (attrition bias)	Low risk	The reported attrition rate is 4 percent, with no systematic differences between groups.
Selective reporting (reporting bias)	Low risk	The study reports all primary and secondary outcomes.
Other bias	High risk	Self-reported data collected immediately and two weeks after exposure. The former probably primed study participants' responses for the latter, which is used in this meta-analysis.

Solomon 1988

Methods	RCT	
Outcomes	Sexual behavior (Notification of sex partners, return for test of cure within required two weeks)	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear how patients were randomized.
Allocation concealment (selection bias)	Low risk	Probably took measures to conceal, allocation given the lab-setting of the study.
Blinding of participants and personnel (performance bias)	High risk	The clinicians provided the edutainment intervention on top of the business as usual intervention.
Blinding of outcome assessment (detection bias)	High risk	Self-reported data collected immediately and two weeks after exposure. The former probably primed study participants' responses in the latter, which is used in this meta-analysis.
Incomplete outcome data (attrition bias)	Low risk	The study's attrition rate is 4 percent.

Selective reporting (reporting bias)	Low risk	The study lacked a protocol, though main outcomes seem to be reported.
Other bias	Low risk	No other issues were found.

Vaughn 2000

Methods	Quasi-experimental (Difference-in-difference)	
Outcomes	Sexual behavior (reduction number sexual partners in previous year). Other behavior outcomes were excluded due to important baseline imbalances.	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Control group comes from only one region.
Allocation concealment (selection bias)	Low risk	Centrally allocated by the research team.
Blinding of participants and personnel (performance bias)	Low risk	The non-invasive nature of the study (small number of surveys conducted for a popular radio broadcast) and the region-level randomization minimizes performance bias.
Blinding of outcome assessment (detection bias)	Low risk	Repeated cross sectional minimize survey effects.
Incomplete outcome data (attrition bias)	Low risk	The study is based on annual repeated cross-sectional surveys.
Selective reporting (reporting bias)	High risk	Due to important baseline imbalances, the meta-analysis could only use some of the outcomes.
Other bias	High risk	Important baseline imbalances for several (excluded) outcomes.

Wang 2016

Methods	RCT	
Outcomes	Knowledge (correct condom use) and Attitudes (importance of HIV testing). Behavior outcomes (condom use in last instance of sexual intercourse) excluded as not powered for it.	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Study participants were randomly allocated with computer random number generator.
Allocation concealment (selection bias)	Low risk	Probably took measures to conceal, allocation given the lab-setting of the study.
Blinding of participants and personnel (performance bias)	High risk	Study participants were selected from a broader project that had surveyed these participants.
Blinding of outcome assessment (detection bias)	High risk	Self-reported panel data collected immediately and two weeks after exposure. The former probably primed study participants' responses in the two-week follow up, which is used in this meta-analysis.

Incomplete outcome data (attrition bias)	High risk	Test information suggest attrition rates above 10 percent.
Selective reporting (reporting bias)	Unclear risk	The research protocol was not found, with some results not fully reported.
Other bias	Low risk	No other issues were found.

S4 Table. Effects on Individual Behavior Outcomes

Behavior outcome and Study Name	Sex	Behavior outcome	SMD	s.e.	Lower limit	Upper limit	Z-value	p-value
Age-difference with sexual partner								
Dupas 2011	F	Age-gap	0.24	0.09	0.07	0.41	2.75	0.01
Banerjee 2017	M	Age-gap	0.04	0.05	-0.06	0.13	0.72	0.47
Banerjee 2017	F	Age-gap	0.04	0.05	-0.06	0.14	0.86	0.39
Dupas 2017	F	Age-gap	-0.02	0.04	-0.10	0.06	-0.55	0.58
Number of sexual partners								
Banerjee 2017	F+ M	Number sexual partners	0.13	0.04	0.04	0.21	2.85	0.00
Vaughan 2000	M	Number sexual partners	0.38	0.15	0.08	0.67	2.48	0.01
Vaughan 2000	F	Number sexual partners	0.23	0.11	0.02	0.45	2.16	0.03
Unprotected sex								
Moyer-Guse 2011	F+ M	Safer sex	0.13	0.11	-0.09	0.35	1.13	0.26
Dupas 2011	F	Teenager childbearing	0.06	0.03	0.00	0.12	1.88	0.06
Dupas 2011	F	Condom use last event	0.19	0.12	-0.04	0.41	1.61	0.11
Jones 2013	F	Vaginal episode w/high-risk partner	0.10	0.12	-0.13	0.33	0.86	0.39
Kearney 2015	F	Teenager birth rate	0.07	0.03	0.02	0.13	2.51	0.01
Banerjee 2017	M	Condom use last event	0.06	0.07	-0.07	0.19	0.86	0.39
Banerjee 2017	F	Condom use last event	0.00	0.06	-0.13	0.12	-0.04	0.97
Dupas 2017	F	Currently pregnant	0.09	0.07	-0.05	0.23	1.27	0.20
Dupas 2017	F	Unprotected sex last 12mon	0.11	0.08	-0.04	0.26	1.41	0.16
STI testing and management								
Solomon 1986	M	Return for STI test of cure	0.20	0.07	0.06	0.35	2.76	0.01
Banerjee 2017	F+ M	Tested last 6 mon	0.38	0.09	0.21	0.56	4.25	0.00

M=male, F=female, SMD=standardized mean difference, s.e.= standard error.

S5 Table. Summary of Findings for the Main Comparison

Entertainment-education narratives compared with no intervention or treatment as usual for reducing risky sexual behaviors				
Patient or population: Youth from the general population Settings: Audiences from mass media shows and individuals recruited in schools Intervention: Entertainment-education narratives delivered in a narrative format Comparison: No intervention, treatment as usual				
Outcomes	Illustrative comparative risks* (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Corresponding risk			
	Entertainment-education narratives			
Age-Gap with sexual partners (self-reported) pre-and post-treatment)	Across studies, the number and age of sexual partners was on average 0.06 standard deviations better (with a 95% CI ranging from -0.06 to 0.19) than in the control condition.	11,882 (3 studies)	⊕⊕⊕⊖ moderate ³	No effect
Number of sexual partners (self-reported) pre-and post-treatment)	Across studies, the number and age of sexual partners was on average 0.17 standard deviations better (with a 95% CI ranging from -0.02 to 0.33) than in the control condition.	6,216 (2 studies)	⊕⊕⊖⊖ low ^{1,2}	Small effect
Unprotected sex (self-reported and objective measures) post-treatment)	Across studies, the number and age of sexual partners was on average 0.08 standard deviations better (with a 95% CI ranging from -0.03 to 0.12) than in the control condition.	16,594 (6 studies)	⊕⊕⊕⊖ moderate ²	Small effect
STI testing and management follow up (self-reported and objective measures) post-treatment)	Across studies, the number and age of sexual partners was on average 0.29 standard deviations better (with a 95% CI ranging from -0.11 to 0.46) than in the control condition.	5,902 (2 studies)	⊕⊕⊖⊖ low ^{3,4}	Small effect
Knowledge of HIV and STI transmission (test scores) pre-and post-treatment)	Across studies, test scores were on average 0.67 standard deviations better (with a 95% CI ranging from 0.32 to 1.02) than in the control condition.	2,699 (7 studies)	⊕⊕⊕⊕ high	Large effect in short-term, small effect in long-term
Attitudes towards behavior intentions and people living with HIV/AIDS (self-reported) pre-and post-treatment)	Across studies, the typical scores were on average -0.04 standard deviations worse (with a 95% CI ranging from -0.18 to 0.10) than in the control condition.	7,795 (4 studies)	⊕⊖⊖⊖ very low ^{1,3,4}	Methodological issues are likely driving inconsistent results.
GRADE Working Group grades of evidence: High quality: Further research is very unlikely to change our confidence in the estimate of effect.				

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

Downgraded for:

¹ Limitations in design.

² Indirectness of evidence.

³ Unexplained heterogeneity or inconsistency of results.

⁴ Imprecision of results.

S1 List. MEDLINE search strategy

HIV/AIDS Keywords

1. HIV infections/pc
2. Sexually transmitted diseases/pc
3. Acquired Immunodeficiency Syndrome/pc
4. (HIV or AIDS).ti,ab
5. or 1-4

Evaluation Keywords

6. Program evaluation/
7. Randomized controlled trials/
8. Evaluation studies/
9. Random allocation/
10. Control trial.ti,ab
11. Controlled trial.ti,ab
12. (rct or rcts).ti,ab
13. Evaluation\$.ti,ab
14. Impact\$.ti,ab
15. (effect or effects or effective or effectiveness).ti,ab
16. or 6-15

Intervention Keywords

17. Intervention\$.ti,ab
18. Communication\$.ti,ab
19. Drama\$.ti,ab
20. Edu-tainment.ti,ab
21. Edutainment.ti,ab
22. Entertainment.ti,ab
23. Film\$.ti,ab
24. Mass media\$.ti,ab
25. Novela\$.ti,ab
26. Radio.ti,ab
27. Soap opera\$.ti,ab
28. Soaps.ti,ab
29. Telenovela\$.ti,ab
30. Television.ti,ab
31. TV.ti,ab
32. Youth.ti,ab
33. Young.ti,ab
34. or 17-33

Sexual behavior Keywords

35. Health behavior/
36. Reproductive health/
37. HIV testing/
38. Risk reducing behavior/
39. Contraception behavior/
40. Sexual behavior/
41. Sexual partners/
42. Safe sex/
43. Sexual Concurrency/
44. Unsafe sex/
45. Condom\$/ut

- 46. Contracep\$.ti,ab
- 47. Sex\$.ti,ab
- 48. Stigma ti,ab
- 49. or 35-48
- 50. 5 and 16 and 34 and 49

Limits: Limited to English language

Publication limits: Multi-center study, Corrected and republished article, Evaluation studies, Journal article, Meta-analysis, Randomized controlled trial, Review literature, Technical report, Validation studies.

S2 List. List of 36 studies excluded from the meta-analysis

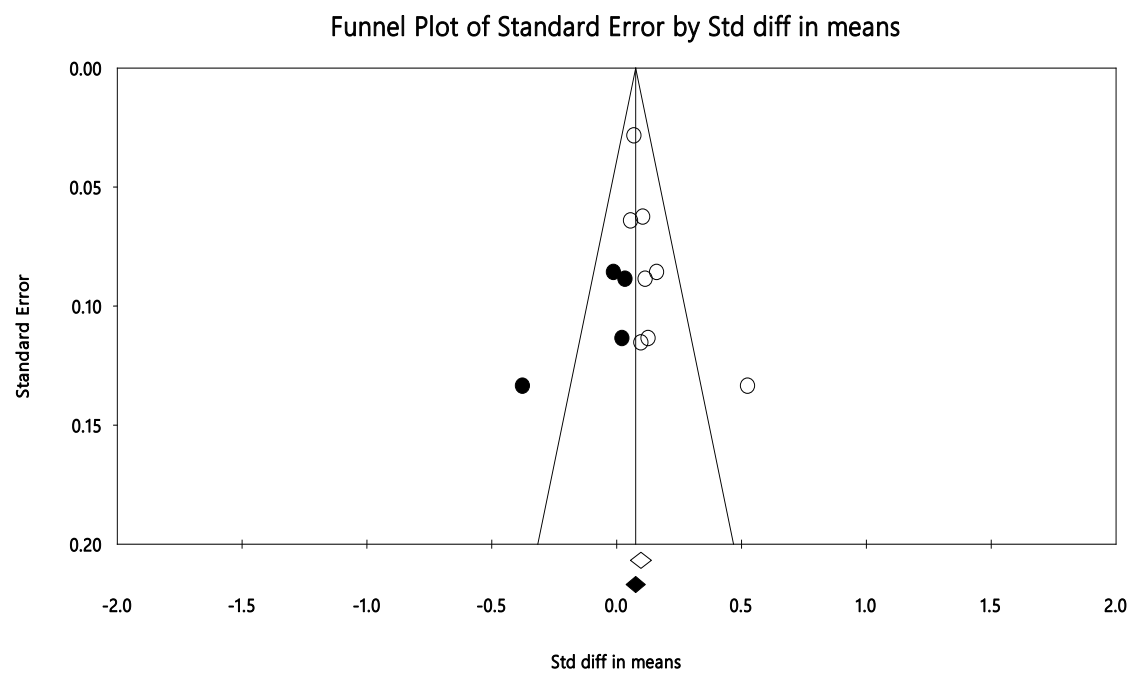
- Anita C, Martino S, Collins R, Elliott M, Berry S, Kanouse D, Miu A. (2008). Does Watching Sex on Television Predict Teen Pregnancy? Findings from a National Longitudinal Survey of Youth, *Pediatrics* 122(5):1047-1054.
- Arroyave J. (2008). Testing the effectiveness of an entertainment-education health-focused soap opera: exposure and post-discussion in Colombian young adults. *Investigación & Desarrollo* 16(2): 232-261.
- Boulay M, Storey JD, Sood S (2002). Indirect exposure to a family planning mass media campaign in Nepal. *Journal of Health Communications*; Oct-Dec 7(5):379-99.
- Brodie M, Foehr U, Rideout V, Baer N, Miller C, Flournoy R, Altman D. (2001). Communicating health information through the entertainment media: A study of the television drama ER lends support to the notion that Americans pick up information while being entertained, *Health Affairs*, 20, 192–199.
- Collins RL, Elliott MN, Berry SH, Kanouse EE, and Hunter SB. (2003). Entertainment television as a healthy sex educator: The impact of condom efficacy information in an episode of Friends. *Pediatrics*, 112, 1115–1121.
- Do M, Kinkaid L. (2006). Impact of an Entertainment-Education Television Drama on Health Knowledge and Behavior in Bangladesh: An Application of Propensity Score Matching. *Journal of Health Communication*; 11:301–325
- Farr AC, Witte K, Jarato K, and Menard T. (2005). The effectiveness of media use in health education: Evaluation of an HIV/AIDS radio campaign in Ethiopia. *Journal of Health Communication*, 10(3), 225-235.
- Geary CW, Burke HM, Castelnuovo L, Neupane S, Sall YB, Wong E, Tucker HT. (2007). MTV's Staying Alive global campaign promoted interpersonal communication about HIV and positive beliefs about HIV prevention. *AIDS Education Prevention*. Feb;19(1):51-67.
- Jadranin Z, Dedic G, Vaughan F, Grillo M, Suljagic V. (2015). The impact of an educational film on promoting knowledge and attitudes toward HIV in soldiers of the Serbian Armed Forces. *Vojnosanit Pregl*;72(7):569-75.
- Jensen R, Oster E. (2009). The Power of TV: Cable Television and Women's Status in India. *The Quarterly Journal of Economics*; MIT Press, vol. 124(3), pages 1057-1094, August.
- Karlyn A. (2001). The impact of a targeted radio campaign to prevent STIs and HIV/AIDS in Mozambique. *AIDS Education Prevention*. Oct; 13(5):438-51.
- Keating J, Meekers D, Adewuyi A (2006). Assessing effects of a media campaign on HIV/AIDS awareness and prevention in Nigeria: results from the VISION Project. *BMC Public Health*. May 3;6: 123.
- Kennedy M, O'Leary A, Beck V, Pollard K, Simpson P (2004). Increases in Calls to the CDC National STD and AIDS Hotline Following AIDS-Related Episodes in a Soap Opera. *Journal of Communication*; 54: 287–301.
- Kuhlmann AKS, Kraft JM., Galavotti C, Creek TL, Mooki M, and Ntuny R. (2008). Radio role models for the prevention of mother-to-child transmission of HIV and HIV testing among pregnant women in Botswana. *Health Promotion International*, 23(3), 260-268.
- La Ferrara E, Chong A, Duryea S (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*; 4 (4), 1-31.
- Lapinski M, Nwulu P (2008). Can a Short Film Impact HIV-Related Risk and Stigma Perceptions?

- Results from an Experiment in Abuja, Nigeria, *Health Communication*, 23:5,403 — 412
- Middlestadt S, Fishbein M, Albarracin D et al (1995). Evaluating the impact of a national AIDS prevention radio campaign in St. Vincent and the Grenadines. *J Applied Social Psychology*; 25: 21–34.
- Mohammed S (2001). Personal communication networks and the effects of an entertainment-education radio soap opera in Tanzania. *Journal of Health Communication*; 6(2), 137-154.
- Moyer-Gus E, Nabi R (2010). Explaining the Effects of Narrative in an Entertainment Television Program: Overcoming Resistance to Persuasion. *Human Communication Research*; 36 p 26–52
- Moyer-Gusé E, Chung A, and Jain P. (2011). Identification with characters and discussion of taboo topics after exposure to an entertainment narrative about sexual health. *Journal of Communication*, 61, 387-406.
- O'Donnell L, Sandoval A, Duran R, O'Donnell CR (1995). The effectiveness of video-based interventions in promoting condom acquisition among STD clinic patients. *Sexually Transmitted Diseases*. Mar-Apr; 22 (2):97-103.
- O'Leary A, Kennedy M, Pappas-DeLuca K, Nkete M, Beck V, Galavotti C (2007). Association between exposure to an HIV story line in The Bold and the Beautiful and HIV-related stigma in Botswana. *AIDS Education Prevention*; Jun 19(3):209-17.
- Pappas-DeLuca KA, Kraft JM, Galavotti C, Warner L, Mooki M, Hastings P, Kilmarx PH. (2008). Entertainment-education radio serial drama and outcomes related to HIV testing in Botswana. *AIDS Education & Prevention*, 20(6), 486-503.
- Peltzer K, Philip S (2004). Evaluation of HIV/AIDS prevention intervention messages on a rural sample of South African youth's knowledge, attitudes, beliefs and behaviours over a period of 15 months. *J Child Adolescent Mental Health*; 16: 93–102.
- Peltzer K, Promtussananon S (2003). Evaluation of Soul City school and mass media life skills education among junior secondary school learners in South Africa. *Social Behavior Perspectives*; 31: 825–34.
- Ramirez-Valles J, Kuhns L, Manjarrez D (2014). Tal Como Somos/Just As We Are: An Educational Film to Reduce Stigma towards Gay and Bisexual Men, Transgender Individuals & Persons Living with HIV/AIDS. *Journal of Health Communications*. Apr; 19(4): 478–492.
- Rogers E, Vaughan P, Swalehe R, Rao N, Svenkerud P, Sood S (1999). Effects of an Entertainment-education Radio Soap Opera on Family Planning Behaviour in Tanzania. *Studies in Family Planning*; Vol. 30, No. 3, pp. 193-211
- Sakha M, Kazerooni P, Lari M, Sayadi M, Azar F, Motazedian N (2013). Effect of an educational intervention on knowledge, attitudes and preventive behaviours related to HIV and sexually transmitted infections in female sex workers in southern Iran: a quasi-experimental study. *International Journal STD AIDS*. Sep; 24(9):727-35
- Shapiro D, Meekers D, Tambashe B (2003). Exposure to the 'SIDA dans la Cite' AIDS prevention television series in Cote d'Ivoire, sexual risk behaviour and condom use. *AIDS Care*; 15: 303–14.
- Silvestre E, Weiner R, Hutchinson P (2015). Behavior change communication and mobile populations: the evaluation of a cross-border HIV/AIDS communication strategy amongst migrants from Swaziland. *AIDS Care*. Sep 2:1-7.
- Smith R, Downs E, Witte K (2007). Drama Theory and Entertainment Education: Exploring the Effects of a Radio Drama on Behavioral Intentions to Limit HIV Transmission in Ethiopia. *Communication Monographs*; Vol. 74, No. 2, pp. 133_153

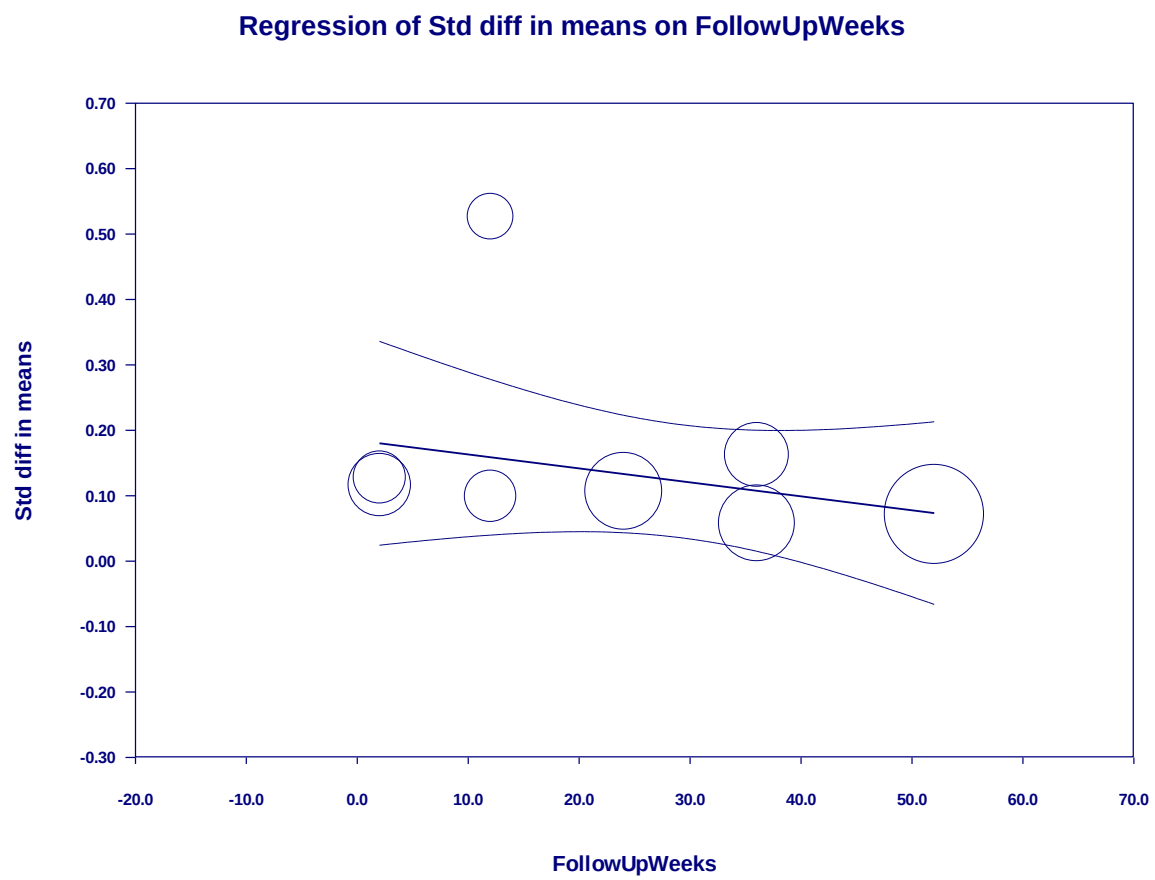
- Storey D, Boulay M, Karki Y, Heckert K, Karmacharya D (1999). Impact of the integrated radio communication project in Nepal, 1994-1997. *Journal of Health Communication*; 4 , 271-294.
- Vaughan PW, Regis A, and St. Catherine E. (2000). Effects of an entertainment-education radio soap opera on family planning and HIV prevention in St. Lucia. *International Family Planning Perspectives*, 148-157.
- Vernon R, Ojeda G, Murad R (1990). Incorporating AIDS prevention activities into family planning organization in Colombia. *Stud Fam Plann*; 21: 335–43.
- Whittier DK, Kennedy MG, St. Lawrence JS, Seeley S, and Beck V. (2005). Embedding health messages into entertainment television: Effect on gay men's response to a syphilis outbreak. *Journal of Health Communication*, 10(3), 251-259.
- Yoder P, Hornik R, Chirwa C (1996). Evaluating the program effects of a radio drama about AIDS in Zambia. *Studies of Family Planning*; 27: 188–203.

S1 Fig. Funnel Plot for Primary Analysis (n=8)

Observed studies in blank, imputed studies in black.



S2 Fig. Meta-regression Effect Size vs Timing of the Follow Up



CHAPTER 4. PAPER 2

Does an educational television series work for the population where it matters most? A post-hoc moderators analysis of youth engaging in high-risk sexual behaviors in Nigeria

Victor H. OROZCO ¹

1/ Economist, World Bank Development Impact Evaluation unit, World Bank Research Group. The World Bank. 1818 H Street, NW Washington, DC 20433 USA.
vorozco@worldbank.org

Abstract

Background: Promoting safer sexual practices among high-risk individuals is key in curbing the HIV epidemic. MTV Shuga is an educational television series aimed at increasing HIV testing uptake and decreasing risky sexual behaviors among African youth. The main paper of the Shuga trial found it was effective for the average study participant. This paper's primary analysis measures Shuga's effectiveness among high-risk individuals ($n=2,073$), broken down by gender. The secondary analysis studies the moderating effects of religion and television-watching habits for this high-risk population.

Methods: This study used panel data from a cluster randomized controlled trial conducted in southwest Nigeria between 2014 and 2016. Nigeria has a generalized HIV epidemic. Data was analyzed from 18-25-year-olds from the urban general population that at baseline reported at least one of the following HIV-risk factors: had unprotected sex in their last sexual encounter, engaged in sexual concurrency, had a sexual partner older/younger by ten years, or reported having symptoms of a sexually transmitted infection (STI). Average treatment effects on knowledge, attitudes, and high-risk sexual behaviors were obtained using an ANCOVA specification. The secondary analysis was operationalized through the interaction of treatment with the moderator variables. Because the study was not designed to study heterogeneous effects for high-risk youth and the mentioned moderators, and because the analysis was not pre-specified, study results should be read as hypothesis-generating rather than hypothesis-testing evidence.

Results: Intent-to-treat estimates suggest that MTV Shuga increased the "short-term" indicator for HIV testing (0 to 8 months after program exposure, self-reported) by 3.3 percentage points (95%CI=0.7 to 5.9), an increase of 32 percent with respect to the control

mean; and the “medium-term” indicator (8 to 10 months, objective) by 4.8 percentage points (95% CI=1.4 to 8.1), an increase of 184 percent . The latter was driven by an increase of 5.4 percentage points (95% CI=0.2 to 8.7) among males, an increase by a factor of five with respect to the control mean of 1.1 percent. The intervention also decreased the average number of current sexual partners of males by -0.14 (95% CI=-0.30 to -0.02), a 10.7 percent reduction. Statistical increases are observed for knowledge about HIV transmission (0.14 standard deviations, 95%CI=0.02 to 0.25) and decreases in stigma and negative attitudes towards HIV-positive people (0.11 standard deviations, 95% CI=0.01 to 0.21). For the combined and gender sub samples, however, no effects were found on chlamydia infections and condom use. The intervention had no effects on inter-generational sex among females. The study did not find evidence that religion or television-viewership moderated program impacts for either gender.

Conclusion: This study suggests that mass media entertainment can be an effective tool in promoting safer sexual practices among high-risk individuals of the general urban youth population. Complementary interventions aimed at economically empowering women may be needed to reduce inter-generational sex. The lack of heterogenous effects for the moderator variables of religion and television-watching habits suggest the potential “transportability” of entertainment-education across different contexts.

INTRODUCTION

In 2017, an estimated 1.2 million people in sub-Saharan Africa became infected with HIV, with more than one-third of new infections occurring in youth between the ages of 15 and 24 (UNAIDS 2018). Systematic reviews of HIV behavior change programs in low- and middle-income countries (Fonner et al. 2014, McCoy et al. 2010), including Sub-Saharan Africa (Harrison et al. 2010, Michielsen et al. 2010, Mavedzenge et al. 2011, Krishnaratne et al. 2016), show these interventions generally improve knowledge and awareness, but have limited impacts on reducing risky sexual behaviors. Systematic reviews for interventions that target individuals at high-risk of HIV infection find mixed evidence for this population. However, most reviews are based on resource-intensive interventions conducted in the US (Abad et al. 2015; Herbst et al. 2007), which limits their relevance for low-resource settings.

Entertainment-education is the purposeful use of mass media programming aimed at educating and improving attitudes and behaviors with development objectives (Singhal and Rogers 1999). Although the promotion of safer sexual behaviors is one of the main applications of entertainment-education (Pappas-DeLuca et al. 2008), educational messages represent a very small proportion of overall media content (Hust, Brown Engle 2008)⁴². Guided by established psychology theories, entertainment-education commonly uses a narrative format (e.g., television dramas)⁴³.

⁴² A quantitative analysis of sexual content of television, magazines, music and movies popular among teenagers in the US shows that less than one half of 1 percent of the content is educational, concluding that significant changes in content would be required for the media to promote safer sexual practices among youth.

⁴³ Social Cognitive Theory (Bandura 1977, 2002), one of the main psychological theories underpinning entertainment-education narratives, states that program characters may serve as role models; and as such, may help improve audiences' self-efficacious beliefs (Bandura 1977). Other theories emphasize entertainment-education's potential to reduce counter-arguing (Benoit 1987; Bensley & Wu 1991; Hall &

MTV Shuga is a television drama produced by MTV Staying Alive Foundation with leading producers and actors from Sub Sahara Africa. It intends to educate African youth about HIV risks, reduce stigma against HIV-positive people, and promote safer sexual behaviors. The Shuga trial studied the third season, composed by eight episodes that lasted a total of three hours. Banerjee, La Ferrara and Orozco (2018) evidence of program effectiveness for the average study participant, which was to some extent representative of the general youth population in southwest Nigeria.

This paper explores program effectiveness on a subsample of high-risk youth, a key population for policymakers. Considering their potential to infect with HIV a larger number of sexual partners, promoting increased HIV testing and reducing risky sex among high-risk individuals has the potential of curbing HIV infections in the broader population (Padian et al. 2011; Jones et al. 2014). To the best of the author's knowledge, there are no field experiments that measure the effectiveness of HIV entertainment-education on high-risk youth. This moderation analysis uses data from the MTV Shuga trial, though it should be noted that its findings should be interpreted as hypothesis-generating and not hypothesis-testing (Spieth et al. 2016) considering its power limitations and its post-hoc nature⁴⁴.

The paper's primary analysis (*subgroup analysis*) intends to explore the following two questions. Did MTV Shuga improve HIV-related outcomes for high-risk youth? Did it separately work for high-risk females and high-risk males? To assess the consistency of

Bracken 2011; Sood 2002) and their potential to positively influence audiences' perceptions of prevalent social norms (Singhal et al. 2004; La Ferrara 2016).

44 The trial's sampling strategy was not designed and powered to measure effects for high-risk youth (which represents 40 percent of the overall sample, with moderator subgroups representing only a quarter of the sample). Moreover, the subgroup analysis is post-hoc (i.e., not pre-specified), so study findings cannot be externally validated with a pre-analysis plan.

program effects across large audience segments, key for HIV targeting efforts (UNAIDS 2016), as secondary analysis the paper explores the potential moderating effects of religion and television-watching habits (*effect-modification analysis*). It was theorized that impacts would be lower for Muslim viewers as Shuga's messages of safe sex rather than abstinence adhere more to Western values. Observational research suggests that changing normative beliefs is greatly contingent on new information not substantially departing from an individual's prior beliefs (Bursztyn, Egorov and Fiorin 2017; Bursztyn, Gonzalez, and Yanagizawa-Drott 2018). Muslim respondents in our sample held more conservative beliefs so Shuga messages may potentially be less effective⁴⁵. Similarly, it was hypothesized that effects would be lower for individuals that watched television on a daily basis, due to theoretical saturation effects of media consumption (Heather et al. 2008). To isolate their moderating influence from gender effects, the effect-modification analysis for religion and television-watching levels is separately conducted for the female and male samples.

To address concerns of social desirability bias (Brener et al. 2003; Minnis et al. 2009), and to the best of the author's knowledge, this is the first study of an HIV entertainment-education intervention that complements self-reported data with objective measures of behavior change.

METHODS

Study Design

⁴⁵ Muslims were less likely to be sexually active, had more negative attitudes towards HIV-positive people, and were more likely to justify a husband hitting or beating his wife if she went out without telling him or if she argued with him. These differences were statistically significant.

The Shuga trial followed a cluster randomized control design. The trial involved community film screenings in 80 locations, defined as two-mile circles centered around a community center (often a school) where screenings took place over weekends ⁴⁶. To focus on the effectiveness of the Shuga *content*, a placebo rather than a pure control was used to “net-out” the effects of the intervention’s delivery mechanism (i.e., community screenings). Though television remains a popular media outlet in Nigeria, urban youth are increasingly watching audiovisual programs online⁴⁷.

Study Location and Participants

The trial took place between 2014 and 2016 in seven cities in southwest Nigeria. This region has an adult HIV prevalence rate of 2.8 percent, below the national average of 3.1 percent (Nigeria Global AIDS Response Report 2017). Thus, the HIV epidemic is *generalized*, defined by UNAIDS (2005 and 2007) as HIV prevalence rates exceeding 1 percent of the general population. Economic and gender power imbalances are prevalent in Nigeria, with survey research in Sub-Saharan Africa indicating power imbalances can adversely affect women’s ability to negotiate safer sex (Langen 2005; Sia et al. 2016)⁴⁸.

⁴⁶ To minimize potential contamination between treatment and control locations, community centers were geographically separated through buffer zones. The trial invited individuals living near these centers to attend three entertainment film screenings (i.e., to focus on the effectiveness of the television drama, the invitation and the intervention did not include further HIV information). Screening zero was the same for both treatment and control groups. This screening was used to select individuals interested in entertainment programs (which is also a more relevant population given the intervention’s media format) and to potentially reduce study participation and attrition rates. Screenings one and two differed by treatment assignment. While treatment locations screened MTV Shuga, control locations screened Gidi-up, a placebo television series with a similar storyline to Shuga’s (i.e., young adults pursuing success and independence in Lagos) that lacked educational HIV messages. To avoid potentially inflating results through the placebo having negative effects on sexual outcomes, the control episodes were selected to minimize storylines heavy with sexual plots.

⁴⁷ According to a marketing study (Consumer Lab Ericsson 2015) in urban locations of Nigeria, 57 percent of 18-24-year-old youth wanted to see television programs at their leisure, with two-thirds of the viewing hours consumed through laptops, mobiles and tablets.

⁴⁸ Nigeria ranks 122 out of 144 countries in the 2017 Global Gender Gap (World Economic Forum, 2017). Population surveys in Sub Sahara Africa show that 15-24-year-old women consistently have higher HIV prevalence rates than their male peers (Joint United Nations Programme on HIV/AIDS 2017).

Study participants were between the ages of 18 and 25. The sample was stratified by gender (50 percent) and oversampled youth in relationships (two-thirds). It was designed to be representative of the general urban-youth population; with the sample being similar in socio-demographic characteristics to a population-level sample of urban southwest Nigeria (NDHS 2008)⁴⁹.

In a post-hoc manner, this study defined as high-risk individuals baseline respondents who reported at least one of the following four HIV-risk factors. In the last six months, (i) had unprotected sex in their last sexual encounter; (ii) engaged in sexual partner concurrency (i.e., simultaneously had two or more sexual partners); (iii) engaged in inter-generational sex, defined as age-disparate relationships with 10 years or more age difference following Balkus et al. (2015) and Schaefer et al. (2017); and (iv) reported symptoms for a sexually transmitted infection (STI). The selection of these high-risk behaviors was informed by considerations relevant to the HIV epidemic in Sub-Saharan Africa⁵⁰. Interestingly, high-

⁴⁹ Other than age, no exclusion criteria were used. Listing data comparing screening attendees versus non-attendees suggests that study participants were greatly representative of invited youth. Table A1 compares the trial's sample versus a population-level sample of urban southwest Nigeria (NDHS 2008). The Shuga sample was similar to the region's averages in terms of religion (one-third Muslim and two-thirds Christian), high levels of schooling (around 11 years of education, twice the national average) and television-ownership rates (nine in ten households). The Shuga sample differed to the region's averages in terms of HIV attitudes and risky sexual behaviors, though differences were not systematically higher or lower.

⁵⁰ Selected HIV risk factors: (i) *Had unprotected sex in their last sexual encounter*. Unprotected sex remains the main source of HIV transmission among youth in Nigeria and Sub-Saharan Africa (Kharsany et al. 2016; UNAIDS 2016; UNICEF 2017). Unprotected heterosexual sex represents 80 percent of new cases of HIV in Nigeria (National Agency for the Controls of AIDS 2015). (ii) *Engaged in sexual concurrency*. This high-risk behavior increases the likelihood of contracting the virus per sexual event as recently-infected individuals are substantially more contagious than those who have had the virus for over six-eight weeks (Epstein and Morris 2011, Koopman, Simon and Riolo 2005). (iii) *Engaged in inter-generational sex*. The HIV literature commonly classifies age-disparate relationships as inter-generational (≥ 10 years difference); as intra-generational (5–9 years difference); and age-homogeneous (0–4 years difference). Research suggests that inter-generational sex between younger women and older men is an important driver of the HIV epidemic in Sub-Saharan Africa as older men tend to have substantially higher HIV prevalence rates in a context of low condom use (Poulin 2007; Dupas et al. 2011; Toska et al. 2015; Schaefer 2017). (iv) *Reported symptoms for a sexually transmitted infection*. High-risk sexual behaviors that lead to STIs also lead to HIV infection. Biological models suggest that active genital sores from other STIs can plausibly facilitate the transmission of HIV, though research has failed to confirm this hypothesis (Ward and Ronn 2010; Ng et al. 2011).

risk individuals have similar levels of HIV knowledge and attitudes than low-risk individuals⁵¹.

Ethics approval for the Shuga trial was secured from the National Health Research Ethics Committee of Nigeria and the ethics committees of MIT, Bocconi University, and Oxford University. The Shuga trial was registered at the American Economic Association RCT Registry (<https://www.socialscisceregistry.org/trials/509>).

Measures and Outcomes

Outcomes

This post hoc study restricted the number of hypotheses tested (in terms of HIV outcomes and measures *per* outcome) based on the theory of change of MTV Shuga, (i) to minimize inflation of false positives (i.e., Type-I error) due to multiple hypothesis testing (Young and Karr 2011; Westfall and Young 1993); and (ii) to address concerns of selectively reporting outcomes with significant results (Gamble et al. 2017). Five *behavioral* outcomes (HIV testing, chlamydia status, sexual concurrency, unprotected sex, and inter-generational sex) and two *psychosocial* outcomes (knowledge and attitudes towards HIV-positive people) were analyzed. A total of 11 measures were analyzed for the primary analysis.

The follow-up survey was conducted 8 months after the community screenings took place. Follow-up respondents were then provided with phone-credit incentives to attend “temporary clinics” in the same 80 community centers that were used for the

⁵¹ Table A2 compares the samples of high-risk and low-risk individuals (i.e., baseline respondents who reported none of the four HIV-risk factors). The high-risk sample is less likely to attend school and more likely to be Muslim, with no differences in terms of television-watching habits. While both samples are similar in terms of knowledge about HIV transmission and general attitudes towards HIV-positive people, by construction the high-risk sample is more likely to report engaging in sexual concurrency and having unprotected sex.

entertainment screenings. Staffed now with clinicians from the University College Hospital-Ibadan, these temporary clinics collected objective measures for behavior change. Details of outcome measures follow:

HIV testing. This outcome was studied through two indicators, which measure the period prior to the follow-up survey (“short-term”) and two months after it (“medium-term”). Respondents that reported having at least once gotten tested for HIV were asked the last time they got tested. Response options included: less than six months ago, 6-12 months ago, 12-13 months ago, and two or more years ago. A binary variable was constructed for individuals that reported getting tested in the last six months. The “medium-term” indicator was an objective behavior measure. Clinic attendees were provided with information notes with the address of the closest HIV testing centers. The information note had an individual identification code and was directly collected by the research team at the HIV testing centers within two months of the follow-up survey⁵².

Sexual partner concurrency. This outcome was defined as the percentage of respondents who concurrently had two (or more) sexual partners. It was measured through two questions. Survey respondents were first asked *if they had engaged in this practice* in the last six months (proportion outcome). Respondents were then asked the *number of sexual partners* they had at the moment of the follow-up data collection (quantitative outcome).

Unprotected sex. This outcome was measured in two ways. First, survey respondents were asked if they used a condom in their last sexual encounter. The

⁵² Following Baird et al. (2012), study participants were not offered HIV testing at baseline as this would have likely changed their sexual behaviors and reduced the number of individuals testing for HIV at follow-up, a primary study outcome.

question was repeated for the last sexual encounter with their main and other sexual partners (the latter required the respondent reporting at least two sexual partners). Second, and similar to Cho et al. (2013), attendees to the temporary clinics were offered either condoms packs or their financial equivalent. It was theorized that with respect to the controls, treated individuals would be more likely to choose the condoms packs.

Inter-generational sex. Survey respondents were asked the age of their last three sexual partners during the six-month period. An inter-generational binary indicator was constructed for individuals that reported having a sexual partner 10-years older or younger.

Infected with a sexually transmitted infection (STI). Clinic attendees provided urine samples to test for chlamydia based on WHO (2013) STI testing guidelines related to informed consent, counselling and confidentiality. Rapid tests included a qualitative, lateral flow immunoassay to detect the chlamydia antigen⁵³.

Knowledge about HIV transmission index. This index included 14 items and aimed to capture a respondent's awareness and knowledge about HIV transmission mechanisms, the existence of antiretroviral drugs, and knowledge about the adequate timing for HIV testing. Response options were open-ended (e.g., name ways to contract HIV) and binary (can a mother transmit the virus to her child during pregnancy? Yes/No).

Stigma and Attitudes towards HIV-positive people index. This index included 7 items and intended to capture respondents' inclination to potentially reveal their status and general attitudes towards HIV-positive people. Response options were either binary (e.g., Do you think a young boy with HIV should be allowed to play football with other

⁵³ Prior to the Shuga trial, researchers conducted a prevalence STI study near study sites. That study found that chlamydia was the most common STI in the study population and, therefore, most likely to increase the statistical power of the biomarker outcome.

boys? Y/N) or based on a 4-point scale (Strongly agree, Agree, Disagree, Strongly disagree).

Practically all behavioral measures were binary, thus program impacts are reported in changes in percentage points. The indexes for knowledge and attitudes were coded in a way where answers consistent with Shuga's objectives translated into higher values. For their construction, principal component analysis for a family of outcomes was used (Anderson 1963, Jackson 2003)⁵⁴. To facilitate interpretation of impacts on the two outcome indexes, results are reported in terms of standard deviations⁵⁵.

Moderators and Measure Reliability

A binary variable was constructed for participants who reported at baseline at least one of the mentioned four HIV-risk factors. The religion moderator variable was binary (Muslim=1, otherwise=0). Practically all non-Muslims were Christians in our sample as both categories covered over 99.5 percent of study participants. Responses were categorical for television-watching habits (Do you watch television almost every day, at

⁵⁴ The Cronbach alphas (Cronbach 1951) for the indexes of HIV knowledge and attitudes are 0.72 and 0.47 respectively. Cronbach's alpha measures reliability or internal consistency, or the average inter-correlation among a set of items. A commonly accepted rule is to consider Cronbach alphas above 0.70 acceptable (Kline 2000; George and Mallery 2003). Low alpha values suggest that the index included items that were not narrowly related (Streiner 2003). This was the objective for these indexes, which intended to be comprehensive measures of several key HIV messages, including those specifically discussed in the three-hour drama. Thus, the items used to construct these indexes may not be reliable measures for other research projects. The use of variables with low internal consistency runs the risk that clinically and statistically significant effects may be less likely to be observed (Moonseong, Namhee and Myles 2015). Breaking the Attitudes index into sub-indices did not qualitatively change internal consistency results.

⁵⁵ Impact effects were divided by standard deviations derived from the control group at follow-up. An increase of two standard deviations, for example, would move someone from having average knowledge to being in the top 5 percent of the group.

least once a week, less than once a week or not at all?). A binary variable was created for individuals that reported watching television every day (daily viewer=1, otherwise=0)⁵⁶.

Most survey scales have been used and validated with Nigerian urban youth (Fatusi and Blum 2008; Odu and Foluso 2008). Most of the HIV knowledge, attitudes and sexual behavior items were derived from the 2008 Nigeria Demographic Health Survey. This population-level survey has been repeatedly administered to the study population since 1990. A small number of knowledge and attitudes items included messages from MTV Shuga (e.g., the need to get tested again after three months of the potential HIV exposure). See Appendix 3 for scripts and main HIV/AIDS messages of MTV Shuga. A series of survey pilots were conducted prior to the baseline to increase response variation and validate the psychosocial and behavior items (Presser et al. 2004). Surveys were conducted in private spaces and enumerators were the same gender of the respondent.

Study Attrition

None of the study 80 locations were dropped from the analysis. The broader Shuga trial reported low and unsystematic attrition rates of the experimental groups⁵⁷. To ensure that attrition did not invalidate the randomization strategy for the high-risk sample, Table A3 shows regression results for the decision to participate in the follow-up survey (Panel A) and to attend the temporary clinics (Panel B) by treatment assignment. Attrition rates

⁵⁶ The choice for daily-viewers divided the sample near its median, with 45 percent of respondents reported watching television every day. Median splits are commonly done in experiment studies as they offer an “impartial” data partition and a way to reduce statistical power loss (Iacobucci et al. 2015).

⁵⁷ At baseline, 5,166 study participants agreed to be part of the Shuga trial. The follow-up survey managed to track and interview 4,986 baseline respondents, which corresponds to an attrition rate of 3.5 percent. At the time of the follow-up, a quarter of the baseline respondents had moved outside the study cities, mostly to attend university. While the survey firm interviewed these individuals in person, these respondents were not expected to attend the temporary clinics (23.1 percent attrition rate). However, out of the 3,901 respondents that remained living in the study cities and that were expected (and financially incentivized) to attend the temporary clinics, 3,832 attended. This represented an attrition rate of 1.8 percent.

are similar to those observed in the broader trial (Banerjee, La Ferrara and Orozco 2018). The attrition rate for the follow-up survey was under 4.4 percent, with no evidence of differential attrition rates by treatment assignment. Thus, different sample compositions are unlikely to influence the study's self-reported results collected by the follow up survey. The attrition rate for temporary clinics was around 23 percent, with no evidence of differential attrition rates by treatment assignment. As seen in Panel B, the baseline outcomes of clinic attendees don't systematically differ from those of non-attendees. Taken together, we find no evidence that the high-risk sample is affected by attrition bias (Dumville, Torgerson and Hewitt 2006).

Measuring Average Treatment Effects and Differential Effects

The subgroup analysis benefits from (i) the Shuga trial being stratified and powered to detect effects by gender; and for the high-risk sample, observed (ii) pre-treatment balance (see Table 1) and (iii) low and unsystematic attrition rates between the treatment and control groups (see Table A3). Thus, point estimates of the subgroup analysis should *approximate* those of hypothesis-testing evidence.

Statistical Analysis

The subgroup analysis was conducted separately for the samples of high-risk females (n=1,029), high-risk males (n=1,044) and their combined samples (n=2,073). Average treatment effects were obtained using an ANCOVA specification controlling for the baseline values of the outcomes of interest (when available) and a series of controls

measured at baseline to improve the efficiency of treatment estimators⁵⁸. Equation 1 shows the ANCOVA specification used in the primary analysis.

$$(1) \quad Y_{ilc1} = \delta Y_{ilc0} + \beta T_{ilc0} + \mathbf{X}'_{ilc0} + \Phi_c + \varepsilon_{ilc1}$$

where: Y_{ilc1} indicates the outcome for individual i who lives in location l within city c during the follow-up (time 1); Y_{ilc0} is the baseline value of the outcome of interest; T_{ilc0} represents a dummy indicator equal to 1 if the individual was assigned to the Shuga treatment and 0 if assigned to the control; $\mathbf{X}'_{ilc0}$ is a vector of controls measured at baseline; and Φ_c are city fixed effects (7 towns). The standard errors were clustered at the screening-location level (80 clusters).

Modeling heterogeneity involves specifying in the regression how moderator variables contribute to higher or lower treatment effects. For the effect-modification analysis, the moderator variables and their interactions with the treatment variable were added to the above specification (Gerber and Green 2012). For interaction terms that showed statistical significance, post-estimation F-tests were conducted to assess whether average treatment effects existed for the specific subgroup (e.g., Muslim females).

The objective measures for HIV testing, condom demand and chlamydia were only collected at the follow-up stage. Given the experimental design of the Shuga trial, the lack of their baseline values should not affect the causal interpretation of both outcomes as both the treatment and control groups are in expectations identical (Bruhn and McKenzie 2009; Todd 2014). This is further confirmed by observed balance for the pre-treatment

⁵⁸ Selected controls were theoretically associated with outcomes and included the respondent's religion, age, currently enrolled in school, number of years of formal education, spoke English as primary or secondary language, Yoruba is spoken as the main language at home, was not living with parents at the moment of the data collection, number of household members, dwelling where the respondent lived was owned or not, and a wealth index constructed from house characteristics and durable goods.

values of observable characteristics and outcomes (Table 1). Ordinary least squares and probit regressions were used for continuous and binary outcomes respectively. All analyses were done by intention to treat (ITT)⁵⁹.

The analysis reports effect sizes, significant results at the 5 percent level, 95% confidence intervals and p-values. All p-values were two-sided. As advised by the American Psychological Association (Wilkinson 1999), results are reported in unstandardized measures given that the units of measurement for the behavioral outcomes are meaningful on a practical level (e.g., increase in percentage points of HIV testing).

RESULTS

Randomization Check and Sample Characteristics

Table 1 displays pre-intervention characteristics of the study sample. High-risk individuals represented 40.1 percent of the baseline sample⁶⁰. The trial's equal stratification by gender persisted for the high-risk sample (49.6 percent is female). Comparisons between experimental groups provides evidence of pre-treatment balance, lending greater credibility to the causal interpretation for the subgroup analysis (World Bank 2011). Pre-treatment differences between experimental groups were generally very small and not

⁵⁹ ITT analysis replicates better what happens in the “real world”, incorporating individuals’ non-compliance or poor adherence to the program. As a result, ITT estimates provides a lower-bound estimate of program impacts.

⁶⁰ The treatment sample is twice as large as the control sample as the Shuga trial included two cluster-level treatment arms. While one treatment group only watched MTV Shuga, a second treatment group was exposed to a five-minute videoclip that provided information about prevalent HIV social norms. This moderation analysis merges both. For more details about the design of the Shuga trial, see Banerjee, La Ferrara and Orozco (2018) and the Template for Intervention Description and Replication Checklist (TIDieR).

statistically significant, including for the self-reported outcomes of HIV testing, sexual concurrency, condom use and STI symptoms⁶¹.

Comparisons between genders shows that each engages in different risky sexual behaviors at different rates at baseline. Females reported being less likely to use condoms in their last sexual encounter with their main sexual partner (17 vs 35 percent) and with secondary partners (48 vs 61 percent); and reported experiencing STI symptoms at a higher rate than males (18 vs 6 percent). While one-eighth of females reported engaging in inter-generational sex, only three males reported engaging in this behavior⁶². On the other hand, high-risk males were five times more likely to report having engaged in sexual concurrency in the last six months (55 vs 10 percent) and reported over twice as many sexual partners at the time of the baseline survey (1.68 vs 0.77). Despite this higher HIV risk, males were only half as likely to get tested for HIV compared to females (5 vs 9 percent).

⁶¹ No indicators were statistically imbalanced for high-risk males. For high-risk females, only one out of fifteen indicators were statistically imbalanced: treated females were more likely to attend school than control females (Control=0.25, Treatment=0.34, $p=0.04$). The analysis controls for this variable.

⁶² This very low number for high-risk males is due to the young age of the study sample. The ten-year age difference with sexual partners meant that for the average male participant, who was 21, his partner had to be either 11 (or younger) or 31 years of age (or older). As a result, the analysis drops program impacts on inter-generational sex for males. For a similar reason, the analysis drops for females the outcome of condom-use with secondary or "other" sexual partners.

Table 1. Socio-demographic and HIV outcomes, Pre-intervention characteristics

Variables	High-risk Female (N=1,029)			High-risk Male (N=1,044)		
	Control (1)	Treatment (2)	p-value diff. (3)	Control (4)	Treatment (5)	p-value diff. (6)
<u>Socio-demographics</u>						
Age (years)	21.4 (1.3)	21.6 (1.3)	0.24	21.4 (1.3)	21.3 (1.3)	0.64
Currently attending school (Yes=1)	257 (25%)	350 (34%)	0.04	334 (32%)	396 (38%)	0.13
Years of education	11.5 (1.43)	11.6 (1.28)	0.60	11.6 (1.26)	11.6 (1.20)	0.94
Speaks English as primary/secondary language (Yes=1)	916 (89%)	946 (92%)	0.18	939 (90%)	939 (90%)	0.97
Does not have a partner (Yes=1)	72 (7%)	61 (6%)	0.66	83 (8%)	104 (10%)	0.44
Religion (Muslim=1)	360 (35%)	401 (39%)	0.53	511 (49%)	407 (39%)	0.16
<u>HIV outcomes</u>						
Tested for HIV in last 6 months (self-reported)	93 (9%)	93 (9%)	0.93	52 (5%)	52 (5%)	0.9
Number of current sexual partners	0.74 (0.58)	0.80 (0.54)	0.19	1.40 (1.23)	1.41 (1.07)	0.88
Sexual Concurrency last 6 mon (%)	110 (12%)	73 (8%)	0.13	562 (56%)	553 (55%)	0.73
Used condom w main partner last time	147 (17%)	146 (17%)	0.92	264 (36%)	256 (35%)	0.74
Used condom w secondary partner last time	47 (49%)	44 (47%)	0.84	355 (63%)	327 (58%)	0.29
Partner age difference >10 years	104 (13%)	96 (12%)	0.62	na	na	
Experienced STI symptoms in last 6 months	195 (19%)	175 (17%)	0.48	62 (6%)	62 (6%)	0.92
HIV Knowledge index (pca scores)	-0.37 (1.74)	-0.32 (1.76)	0.77	-0.50 (1.47)	-0.30 (1.63)	0.10
HIV Attitudes index (pca scores)	-0.20 (1.32)	-0.27 (1.30)	0.62	-0.27 (1.34)	-0.29 (1.33)	0.87

For continuous variables, the table presents means and standard deviations. For binary variables, the table shows the sample size and the proportion of the variable in percent terms.

Subgroup Analysis

Table 2 presents the intent-to-treat results of the subgroup analysis for the high-risk sample, with its breakdown per gender. Following 2010 CONSORT reporting guidelines (Moher et al. 2010), unadjusted means and standard deviations are presented for both treatment and control groups at follow up. The table also reports the adjusted impact estimates and their respective p-values when controlling for a series of covariates, including the baseline value of the outcome of interest. For practically all outcomes, the unadjusted and adjusted point-estimates are very similar. With the exception of condom use, the *coefficient signs* of all outcomes are in the direction of increased safer sexual practices. As rare outcomes, the absolute effects sizes are small, but their percent change is often large.

In this exploratory analysis, the data suggests that MTV Shuga increased the “short-term” indicator for HIV testing (0 to 8 months after program exposure, self-reported) by 3.3 percentage points (95% CI=0.7 to 5.9), an increase of 32 percent with respect to the control mean. This was driven by an increase of 4.8 percentage points (95% CI=3.1 to 9.1) among females, which represented a 41 percent increase. The intervention had larger effects on the objective HIV testing indicator (8 to 10 months after program exposure). Shuga increased this indicator by 4.8 percentage points (95% CI=1.4 to 8.1), which represented an increase of 184 percent with respect to the control mean of 2.6 percent. This impact was in turn driven by an increase of 5.4 percentage points (95% CI=0.2 to 8.7) among high-risk males, an increase by a factor of five with respect to the control mean of 1.1 percent.

Table 2. Average Treatment Effects for high-risk youth, Gender breakdown

Outcomes	Control (means, SD)	Treatment (mean, SD)	Adjusted difference, percentage points, pca scores (95% CI)	p- value	n
Short-term HIV testing (self-reported)	10.3% (0.30)	13.3% (0.34)	3.3 p.p. (0.7 to 5.9)	0.013	1990
Female	11.6% (0.32)	15.6% (0.36)	4.8 p.p. (3.1 to 9.1)	0.036	968
Male	9.1% (0.29)	11.1% (0.31)	1.8 p.p. (-1.9 to 5.6)	0.330	1022
Medium-term HIV testing (observed)	2.6% (0.16)	7.2% (0.26)	4.8 p.p. (1.4 to 8.1)	0.006	1559
Female	4.3% (0.20)	8.0% (0.27)	3.9 p.p. (-0.0 to 0.08)	0.081	772
Male	1.1% (0.11)	6.5% (0.25)	5.4 p.p. (0.2 to 8.7)	0.002	787
Tested positive for chlamydia (%)	3.2% (0.17)	2.0% (0.14)	-1.2 p.p. (-4.3 to 1.8)	0.418	1557
Female	3.4% (0.18)	2.2% (0.15)	-1.3 p.p. (-4.4 to 1.8)	0.410	770
Male	3.1% (0.17)	1.7% (0.13)	-1.4 p.p. (-4.9 to 2.2)	0.448	787
Number current sexual partners	1.03 (0.92)	0.97 (0.82)	-0.059 (-0.14 to 0.02)	0.142	1988
Female	0.74 (0.51)	0.78 (0.48)	0.02 (-0.05 to 0.09)	0.528	967
Male	1.30 (1.12)	1.15 (1.01)	-0.14 (-0.30 to -0.02)	0.035	1021
Sexual Concurrency last 6 mon (%)	24.2% (0.43)	20.8% (0.41)	-2.7 p.p. (-6.4 to 0.1)	0.142	1798
Female	6.4% (0.24)	5.3% (0.22)	-1.1 p.p. (-0.4 to 0.2)	0.412	866
Male	40.2% (0.49)	35.4% (0.48)	-4.1 p.p. (-10.4 to 2.2)	0.201	932
Chose condoms over money (%)	36.5% (0.48)	36.8% (0.48)	0.2 p.p. (-5.7 to 6.2)	0.946	1990
Female	21.9% (0.41)	26.4% (0.44)	3.6 p.p. (-2.4 to 9.6)	0.238	968
Male	50.3% (0.50)	46.9% (0.50)	-2.7 p.p. (-10.7 to 5.2)	0.497	1022
Used condom w main partner last (%)	40.0% (0.49)	37.3% (0.48)	-0.9 p.p. (-6.1 to 4.2)	0.724	1552
Female	31.8% (0.46)	27.1% (0.44)	-4.7 p.p. (-11.3 to 1.9)	0.160	794
Male	48.5% (0.50)	48.2% (0.50)	4.0 p.p. (-5.1 to 13.1)	0.384	758
Used condom other partner last (%) a/	70.8% (0.45)	64.5% (0.48)	-0.5 p.p. (-10.8 to 9.9)	0.925	375
Partner age difference >10 years (%) b/	10.6% (0.31)	11.1% (0.31)	1.1 p.p. (-3.2 to 5.6)	0.605	862
HIV Knowledge index (pca scores) c/	0.083 (1.57)	0.377 (1.67)	0.214 (0.039 to 0.388)	0.017	1889
Female	0.186 (1.56)	0.524 (1.72)	0.290 (0.039 to 0.541)	0.024	935
Male	-0.020 (1.58)	0.234 (1.61)	0.156 (-0.089 to 0.402)	0.210	954
HIV Attitudes index (pca scores) c/	0.056 (1.51)	0.231 (1.53)	0.164 (0.011 to 0.317)	0.036	1990
Female	0.009 (1.50)	0.180 (1.49)	0.187 (0.004 to 0.369)	0.045	968
Male	0.099 (1.52)	0.279 (1.56)	0.143 (-0.101 to 0.388)	0.247	1022

Standard errors adjusted for clustering at the screening location (80 clusters). Increases in indexes correspond to outcomes more consistent with the messages of Shuga. a/ Males only, b/ Females only, c/ Effects reported in the text are in standard deviations, which are obtained by dividing program impacts in pca scores by the standard deviation of the control group. Effects in the converted scale are: *Knowledge*: All 0.14 SD (0.02-0.25 SD), Female 0.19 SD (0.03 to 0.35 SD), Male 0.10 SD (-0.06 to 0.25 SD). *Attitudes*: All 0.11 SD (0.01-0.21 SD), female 0.12 SD (0.00 to 0.25), Male 0.09 SD (-0.07 to 0.26 SD).

The data suggests a reduction of sexual partner concurrency among males. The average number of current sexual partners (quantitative outcome) decreased by -0.14 (95% CI=-0.30 to -0.02), a 10.7 percent reduction. Sexual concurrency did not decrease among females, potentially due to floor effects (only 6.4 percent of control females, compared to 40.2 percent of control males, reported engaging in sexual concurrency).

The intervention had impacts on both psychosocial outcomes. Statistical increases are observed for the indexes of knowledge about HIV transmission (0.14 standard deviations, 95% CI=0.02 to 0.25) and decreases in stigma and negative attitudes towards HIV-positive people (0.11 standard deviations, 95% CI=0.01 to 0.21). Impacts on females drove these results. The program increased females' knowledge and attitudes by 0.19 standard deviations (95% CI=0.03 to 0.35) and 0.12 standard deviations (95% CI=0.00 to 0.25), respectively. For males, no statistically significant differences were found for either psychosocial outcome.

The intervention did not impact all outcomes of interest. For the combined and gender sub samples, no effects were found on chlamydia infections and condom use. Among females, the intervention had no effects on inter-generational sex.

Treatment effects are similar in magnitude to those reported for the average study participant (Banerjee, La Ferrara and Orozco 2018), suggesting consistency of program effects between average study participants and high-risk youth. Effect sizes were similar in terms of HIV knowledge, attitudes towards HIV-positive people, and decreased sexual concurrency. Similarly, the intervention had no impacts on condom use and inter-generational sex for both populations. However, important differences emerged in HIV testing and STI infections. A positive finding is that effect sizes were larger for HIV testing among high-risk youth for both the short-term outcome (3.3 versus 2.5 percentage points) and the medium-term outcome

(4.8 versus 3.1 percentage points). On the other hand, while the main paper reported a decrease of chlamydia infections among females by 1.7 percentage points (a 55 percent decrease with respect to the control rate), the decrease of 1.3 percentage points among high-risk females was not statistically significant. Although the latter may potentially be explained by power limitations, we cannot discard the lack of effects among high-risk females⁶³.

Effect-modification Analysis

Table 3 presents the secondary analysis per gender for the hypothesized moderators of religion and television-watching habits. It reports the means and standard deviations of the control group, the coefficients and the p-values of the treatment*moderator interaction term (i.e., this coefficient measures the outcome difference between Muslims and Christians, and between television daily viewers and non-daily viewers).

For both genders, the effect modification analysis suggests no differential effects existed between Muslim and Christian participants. Only one out of 20 differential outcomes were statistically significant, so by chance this may be a spurious result⁶⁴.

The data also suggests an overall lack of differential effects by television-watching habits, with only two out of 20 differential outcomes being statistically significant. No differential effects were observed for treated male daily viewers with respect to male non-daily viewers. For the female subsample, we observe two significant differences though these go in opposite directions.

⁶³ To detect the 1.3 percentage point decrease required a ten-fold increase of the sample size.

⁶⁴ Muslim females became less likely to use condoms with their main partners compared to Christian females. The joint-test for the treatment and interaction coefficients suggested condom use decreased among Muslim females (Prob> F=0.001).

Table 3. Moderating Effects of Religion and Television-Consumption Levels

Outcomes	Control (mean, SD)	Religion d/ percentage points, means, pca scores	p- value	TV watching habits e/ percentage points, means, pca scores	p- value	n
Short-term HIV testing (self-reported)						
Female	11.6% (0.32)	2.1 p.p.	0.50	4.4 p.p.	0.100	1029
Male	9.1% (0.29)	-0.4 p.p.	0.86	-1.1 p.p.	0.630	1044
Medium-term HIV testing (objective)						
Female	4.3% (0.20)	0.9 p.p.	0.80	3.2 p.p.	0.200	772
Male	1.1% (0.11)	-0.7 p.p.	0.71	0.1 p.p.	0.960	787
Tested positive for chlamydia (%)						
Female	3.4% (0.18)	0.3 p.p.	0.79	-0.04 p.p.	0.071	770
Male	3.1% (0.17)	-0.1 p.p.	0.90	-0.3 p.p.	0.802	787
Number of current sexual partners						
Female	0.74 (0.51)	0.03	0.41	-0.00	0.886	967
Male	1.30 (1.12)	0.14	0.08	-0.03	0.665	1021
Sexual Concurrency last 6 mon (%)						
Female	6.4% (0.24)	0.2 p.p.	0.92	1.6 p.p.	0.553	866
Male	40.2% (0.49)	2.5 p.p.	0.64	1.9 p.p.	0.719	932
Chose condoms over money (%)						
Female	21.9% (0.41)	-0.9 p.p.	0.84	-5.4 p.p.	0.125	968
Male	50.3% (0.50)	-2.3 p.p.	0.60	-2.0 p.p.	0.587	1022
Used condom main partner last (%)						
Female	31.8% (0.46)	-0.14 p.p.	0.02	5.1 p.p.	0.441	794
Male	48.5% (0.50)	-8.7 p.p.	0.21	8.8 p.p.	0.135	758
Used condom second partner last a/	70.8% (0.45)	-0.3 p.p.	0.98	-6.2 p.p.	0.579	375
Partner age difference >10 years b/	10.6% (0.31)	-1.3 p.p.	0.56	5.0 p.p.	0.026	862
HIV Knowledge index c/						
Female	0.186 (1.56)	-0.232	0.06	0.146 (0.09 SD)	0.281	935
Male	-0.020 (1.58)	-0.093	0.53	0.256 (0.16 SD)	0.098	954
HIV Attitudes index c/						
Female	0.009 (1.50)	-0.047	0.64	0.322 (0.21 SD)	0.007	968
Male	0.099 (1.52)	-0.007	0.95	0.06 (0.04 SD)	0.676	1022

Standard errors adjusted for clustering at the screening location (80 clusters). Increases in indexes correspond to outcomes more consistent with the messages of Shuga. a/ Males only, b/ Females only, c/ Interaction effect reported in pca scores and converted in standard deviations of the control group, d/ Religion moderator: Muslim=1, Otherwise=0, e/ Television watching habit: Daily viewer=1, otherwise=0

While female daily viewers had higher gains in terms of HIV attitudes compared to female non-daily viewers (0.21 standard deviations, $p=0.01$); female daily viewers became more likely to report engaging in inter-generational sex than non-daily viewers (5.5 percentage points, $p=0.03$)⁶⁵.

Because coefficients of the moderator interaction terms were small and don't show a general pattern, lack of differential effects and not statistical power limitations (Gerber and Green 2012) are likely to explain potential consistent treatment effects across these different subpopulations (Alosh et al. 2015).

DISCUSSION

To the best of the author's knowledge, this post-hoc analysis is the first to study the effectiveness of an HIV entertainment-education intervention on high-risk individuals using field experimental data. Within the broader HIV literature, this is one of few HIV studies that document larger behavioral effects for males⁶⁶.

The intent-to-treat results of the subgroup analysis suggest that MTV Shuga has small but statistically significant effects on increasing knowledge about HIV transmission; and helped reduce stigma and negative attitudes towards HIV-positive people. Impacts on psychosocial outcomes have the potential to promote safer sexual practices. A recent mediation study

⁶⁵ An increase in more favorable attitudes towards dating "sugar daddies" or older men may partly explain this result. Treated females were 5.7 percentage points (95% CI=0.01-0.11) more likely to agree or strongly agree with the statement: "It is OK for a young girl to date an older married man if he offered to buy her clothes, shoes and take her out".

⁶⁶ Studies that find larger effects for males include Coates et al. (2014) that experimentally studied community-based HIV interventions in four developing countries; and Jones, Orozco and Rascon (2019) that experimentally studied a vocational training that included a soft-skills component targeted at HIV vulnerable youth in Malawi.

(Orozco and Cluver 2019) using data from the Shuga trial finds that increased HIV testing was mediated by increases in knowledge and attitudes.

MTV Shuga increased HIV testing for both the short-term (self-reported) and medium-term (objective) indicators. The increase in the medium-term measure for men represented an increase by a factor of five. This finding combined with an observed decrease in the number of concurrent sexual partners among males is policy-relevant within the epidemiological context of Nigeria and Sub-Saharan Africa⁶⁷. Considering their higher potential to contract and spread HIV, behavioral effects on high-risk males may help prevent new HIV infections at the population level (Padian 2011).

The lack of impacts on condom use for both genders is a disappointing result. The null effects are further confirmed by the lack of effects on behavioral intentions and attitudes towards condom use. Poor access cannot explain these results, as four in five respondents noted they could get a condom within ten minutes. This finding is in contrast with previous studies that have found that HIV entertainment-education increased condom use among male patients of STI clinics in the US (Jones et al. 2013) and among student teenager populations in the US (Kearney and Levine 2015) and Sub-Saharan Africa (Dupas 2011; Dupas et al. 2017). These contrasting findings on *condom use* may potentially be explained by our study participants (i) shifting to safer partners (i.e., learning the HIV-status of sexual partners and a decrease in sexual concurrency) and (ii) by unprotected sex being less costly for our sample, who is older and less likely to be in school.

⁶⁷ Compared to female-controls, male-controls were four times less likely to get tested for HIV and six times more likely to report sexual concurrency. These gender differences are observed in population-level studies as well. Using epidemiological models, Case et al. (2012) estimated that sexual concurrency represented a fifth of new HIV infections in Nigeria for the 2003-2012 period. Population surveys from Sub-Saharan Africa show that compared to women, men are less likely to get tested for HIV and therefore to start treatment (Cohen et al. 2011; UNAIDS 2016).

The intervention did not reduce inter-generational sex among females. This result may not be surprising considering that this non-financial intervention was unlikely to have addressed economic factors, which greatly drive inter-generational and transactional sex (Poulin 2007; Dupas et al. 2011; Toska et al. 2015; Schaefer 2017). This finding is supported by recent experimental research in Malawi, where a vocational training program reduced transactional sex among vulnerable young females (Jones, Orozco and Rascon 2019).

The effect-modification analysis suggests program effects did not differ by audience's religion nor their watching habits, providing indicative evidence regarding the consistency of program effects (Alosh et al. 2015) for large segments of our high-risk sample. MTV Shuga was not less effective in promoting safer sexual practices among Muslim viewers, which held more conservative views than their Christian peers⁶⁸. This finding provides quasi-experimental evidence to support observational research in the US with individuals with strong religious beliefs. That research suggests that mass media can favorably change these audiences' attitudes about traditional gender roles and homosexual rights (Peterson and Donnerwerth 1998) and make them more accepting of scientific evidence (Collins and Sturgill 2013)⁶⁹.

In addition, this moderation study does not find evidence that the program was less effective with individuals that watched television every day. This may suggest media-consumption levels in our sample remain under the theorized media-saturation threshold; with entertainment-education remaining an effective tool. This finding is consistent with

⁶⁸ This finding is relevant for MTV Staying Alive Foundation, the producer of MTV Shuga, given the continental distribution of this program, including in the northern-Muslim parts of Nigeria.

⁶⁹ However, program effectiveness for other cultural factors may be less clear. For example, Murphy et al. (2015) finds that an educational narrative with messages about the importance of vaccinating against the human papillomavirus was more effective in promoting vaccination uptake for the population whose culture was portrayed in the program (Mexican-American versus White and African-American).

observational research in the US (Heather et al. 2008) that finds that greater media exposure of positive messages improved health intentions of audiences.

The study has two important strengths. First, it provides strong evidence that MTV Shuga was effective in promoting safer sexual practices among high-risk youth. The internal validity of study findings for this subsample was confirmed by (i) evidence of pre-intervention balance of the baseline data and (ii) adjusted and unadjusted estimates of program impacts being stable across practically all outcomes of interest (including the objective measures). Both points reduce the likelihood that impact estimates are biased. Second, this paper provides suggestive evidence that MTV Shuga worked across different subpopulations, key for generalizability claims of study findings (Cook and Campbell 1979). This television drama was effective in promoting safer sexual behaviors for both high-risk females and high-risk males (though being more effective for males, partly because they began with higher levels of risky sexual behaviors). Program effectiveness did not vary by participants' religion or television-watching habits. The lack of heterogeneous effects is an important finding considering the potential of mass media to reach large segments of the population.

This study had two important limitations. First, the selection of outcomes and moderator variables was not pre-specified. Although this paper provides a detailed reasoning behind their selection, concerns regarding reporting bias cannot be externally validated with a pre-analysis plan (Chan 2005; Gamble et al. 2017). Second, statistical power limitations likely affected study findings, increasing the likelihood of Type II error: true effects not statistically detected with the smaller sample (the subgroup and the effect-modification analyses respectively represented less than half and a quarter of the sample size for which the trial

was powered for). However, as noted in the above paragraph, impact estimates are unlikely to be biased.

There are a number of knowledge gaps that follow from this study's findings and would benefit from further research. Future experimental research should be conducted in more normal settings, namely, using public television broadcasts or online platforms (in both saturated and non-saturated media markets). In contrast to one-off or infrequent behaviors, such as HIV testing, the lack of effects on condom use suggests greater research is needed to understand how to trigger habit formation effectively for consistent behavior change in the long-term. Population surveys in the US (Shafii et al. 2004) and South Africa (Hendriksen et al. 2007) suggest habit formation is a necessary condition for promoting consistent condom use among youth. To produce hypothesis-testing evidence for urban high-risk youth, future studies should be designed to measure program effectiveness for this population; as well as other populations at higher risk of contracting and spreading HIV⁷⁰.

CONCLUSION

With the rapid expansion of mass media in developing countries, entertainment-education offers an important opportunity to promote positive behavior change (Moreau 2011; World Bank 2016). This post-hoc study provides evidence that suggests that soap operas may be an effective instrument in promoting HIV testing among high-risk youth, which are likely to have higher HIV prevalence rates (UNAIDS 2014)⁷¹. The study also provides evidence that

⁷⁰ Survey research in Sub-Saharan Africa suggests that HIV prevalence rates of men that have sex with men (Dunkle et al. 2013, Sanders et al. 2013), people who inject drugs (Bowring et al. 2013, Peterson 2013) and sex workers (Baral et al. 2016) are often three times higher than the prevalence rates observed in the general population. In Nigeria, these three groups represent an estimated 3.4 percent of the population, yet they account for 32 percent of new HIV infections (Nigeria Global AIDS Response Report 2017).

⁷¹ HIV+ people are the group central group in the 90-90-90 targets of the UNAIDS Fast Track Strategy (UNAIDS 2014). By 2020, this target intends for 90 percent of people living with HIV to get tested and learn their status; of whom 90 percent are on treatment; of whom 90 percent are virally suppressed. It is estimated that only 38

entertainment-education narratives reduced sexual concurrency among high-risk males. Considering (i) the program's effectiveness across different segments of urban high-risk youth (despite their small size in percentage points); and (ii) the low costs of television and online broadcasts and distribution, mass media entertainment can potentially be a cost-effective tool for preventing HIV infections at the population level (Wakefield, Loken and Hornik 2010). The policy and fiscal advantages of mass media entertainment over more resource-intensive approaches are clearer in countries with generalized HIV epidemics, where UNAIDS (2005 and 2007) recommends targeting all segments of society.

ROLE OF THE FUNDING SOURCE. The sponsor of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

percent of HIV-positive people in Nigeria know their status, below the regional average of 45 percent (Nigeria Global AIDS Response Report 2017).

Table A1. Comparison Shuga Sample with Population-Level Survey (urban areas in Southwest Nigeria)

Variable	Shuga Baseline		DHS - South West	
	N	Mean	N	Mean
<u>Socio-demographics</u>				
Age	5,166	20.6	2,186	21.4
Gender (Female =1)	5,166	47%	2,186	68%
Christian	5,166	63%	2,186	67%
Muslim	5,166	37%	2,186	33%
Years of education	5,166	11.6	2,186	10.6
Household owns a television	5,166	91%	1,475	87%
House has cement walls	5,166	97%	1,475	82%
<u>HIV Attitudes</u>				
If a member of your family got HIV, would you want it to remain a secret?	5,166	68%	1,961	55%
If you knew that a shopkeeper had HIV/AIDS, would you buy fresh vegetables from him/her?	5,166	42%	1,961	55%
People with HIV should be blamed for bringing the disease into the community.	5,166	36%	1,961	64%
<u>Sexual behaviors</u>				
Used condom last time	2,678	51%	1,396	36%
Ever tested for HIV	5,155	26%	1,961	37%
Experienced STI symptoms in the last 6/12* months	5,166	5%	2,186	4%
Number of sexual partners in the last 6/12* months (among those who had started having sex)	2,933	1.24	1,396	1.11

*The DHS used a recall period of 12 months, while the Shuga trial used a recall period of 6 months. The latter was determined by the timing of the follow-up survey, originally planned six months after the intervention was conducted.

Table A2. Pre-intervention comparison of Sample High-risk pool versus Low-risk pool

Variables	High-risk	Low-risk	High-risk	Low-risk
	Mean		N. Obs.	
<u>Socio-demographics</u>				
Gender (female=1)	50%	46%	2,073	3,093
Single (does not have a partner)	8%	33%	2,073	3,093
Age	21.4	20.0	2,073	3,093
Attending school	34%	43%	2,073	3,093
Religion (Muslim=1)	40%	35%	2,073	3,093
Watches TV everyday	54%	52%	2,073	3,093
<u>HIV outcomes</u>				
Tested for HIV last six months (self-reported)	7%	4%	2,073	3,093
Number of current sexual partners	1.09	0.25	2,072	3,088
Sexual concurrency in last 6 months	33%	0.00	1,919	1,014
Used condom w primary partner last time	25%	100%	1,598	740
Used condom w secondary partner last time	58%	97%	659	153

Increasing values of the dependent indexes correspond to outcomes more consistent with the messages of Shuga.

Table A3. Attrition Rates

**Panel A. Attrition rates of Follow-up survey
(Self-reported outcomes)**

	Control	Treatment	Diff. p-value	n
Shuga sample	3.5%	3.5%	0.93	5166
High-risk subsample	4.4%	3.8%	0.63	2073

**Panel B. Attrition rates of Temporary clinics
(Objective outcomes)**

	Control	Treatment	Diff. p-value	n
Shuga sample	22.8%	23.4%	0.70	4986
High-risk subsample	20.7%	22.1%	0.49	1990
Baseline outcomes				
Tested for HIV in last 6 months (self-reported)	6.5%	7.2%	0.58	1990
Sexual Concurrency last 6 mon (%)	36.2%	33.1%	0.30	1845
Used condom w main partner last time	30.8%	24.4%	0.06	1533
Experienced STI symptoms in last 6 months	12.2%	11.3%	0.65	1990

Standard errors adjusted for clustering at the screening center location (80 clusters).

CHAPTER 5. PAPER 3

How Did an Entertainment-Education Intervention Promote HIV Testing? Mediation Analysis of the MTV Shuga Television Series

Victor H. OROZCO^{1*} and Lucie D. CLUVER²

1/* Corresponding co-author. Economist, World Bank Development Impact Evaluation unit, World Bank Research Group. The World Bank. 1818 H Street, NW Washington, DC 20433 USA. vorozco@worldbank.org

2/Professor of Child and Family Social Work, University of Oxford. Department of Social Policy and Intervention, 32 Wellington Square, OX1 2ER. lucie.cluver@spi.ox.ac.uk

Abstract

Background: HIV testing is the gateway to treatment, which benefits the individual and the general population through lower HIV infections. A recent trial showed that the television drama MTV Shuga increased HIV testing among Nigerian urban youth. By decomposing treatment effects, this study investigates the mediating effects of (i) knowledge about HIV transmission, (ii) stigma and general attitudes towards HIV-positive people, and (iii) program transportation experiences on increased HIV testing, in the short-term (0 to 8 months) and the medium-term (8 to 10 months).

Methods: This study used data from a cluster randomized controlled trial conducted in southwest Nigeria between 2014 and 2016. Data was analyzed from 18-25-year-old youth that were sexually active at baseline and for which the trial had data for the outcome and the three mediating variables at follow-up ($n=1,894$). Single mediator models were conducted to identify the average causal mediation effect (Hicks and Tingly 2011). Multiple mediators models (Preacher and Hayes 2008) were conducted to better assess the relative influence of each mediator variable and to assess the robustness of study results.

Results: Despite the program's small effects on HIV testing, which oscillated between 2.6 and 3.2 percentage points across the different mediation models, we are able to detect statistically significant mediating effects. These jointly represented over half of the short-term effects. Gains in knowledge about HIV transmission mediated 0.6 percentage points (95% CI=0.3 to 1.0); gains in more favorable attitudes towards HIV-positive people mediated 0.6 percentage points (95% CI=0.2 to 1.0); and greater program transportation or immersion while watching the television show mediated 0.7 percentage points (95% CI=0.2 to 1.3). However, mediation coefficients substantially decreased in the medium-term (despite

impacts on HIV testing being similar in magnitude), representing a fifth of the short-term estimates. As a result, they were no longer statistically significant. Transportation had the strongest mediation influence on increased HIV testing of the three mediators. While it mediated 22.2 percent of the total effect on the short-term indicator, knowledge and attitudes mediated 14.8 and 18.5 percent, respectively. Its relative mediation influence persisted for the medium-term indicator (6.3 versus 3.1 and 1.3 percent).

Conclusion: The mediating effects of program transportation were at least as influential as the high-level constructs central to reasoned-oriented behavior models. This finding is consistent with the theoretical underpinnings of entertainment-education and with modern cognitive science that indicates that emotions can also play a key role in the “rational” decision-making process, especially among youth. It also has policy implications for HIV information campaigns in Sub-Saharan Africa. Considering that HIV awareness is already high among youth in several sub-Saharan countries, the mediating influence of knowledge gains are likely to further diminish. Thus, effective HIV prevention campaigns need to invest in new formats that while appeal to youth’s rational side, should also aim to reach youth at a deeper and more emotional level (Singhal et al 2004). Considering the diminished role of tested mediators in the medium term, greater research is needed to identify the mediators that were behind the sustained increase of HIV testing.

INTRODUCTION

Storytelling has commonly been used in transmitting prescriptive and normative information in most cultures for thousands of years (Sabido 2002). However, didactive forms of communication have gain ground over narratives in conveying modern health messages (Murphy et al. 2015). Entertainment-education is the purposeful use of mass media programming aimed at educating and improving attitudes and behaviors with development objectives (Singhal and Rogers 1999). A recent meta-analysis of entertainment-education narratives (Orozco, Shen and Cluver 2019) suggests that this communication approach promoted safer sexual behavior among youth in developed and developing countries. This paper intends to provide insights on *how* it works.

A theory-based approach, entertainment-education often aims first to influence two *individual*-level psychosocial mediators⁷² central to reasoned-oriented behavior models (Bandura 2004, Ajzen 2014): *knowledge* (i.e., raising awareness and educating viewers about an issue) and *attitudes* or viewers' preferences for a given action⁷³. Current neuroscience research suggests that emotional and other non-cognitive components also serve as inputs to the "rational" decision-making process (Bosse, Jonker and Treur 2008; Gutnik et al. 2006). Kahneman (2011) argues that intuitive and automatic thinking (System 1) often steers analytical thinking (System 2), especially for routines where conscious and deliberate cognitive thinking would be too burdensome⁷⁴. In contrast to overtly persuasive messages,

⁷² In contrast to structural or macro-social influences such as poverty, healthcare access, or cultural or legal norms (Zimmerman et al. 2007; Mirkuzie et al. 2011; Kaufman et al. 2015).

⁷³ Whereas social cognitive theory (Bandura 2004) argues that behaviors are learned by observing and imitating the behaviors of models; the theory of planned behavior (Ajzen 2014) states that behavioral intentions are determined by an individual's attitudes or beliefs of a behavior as positive or negative.

⁷⁴ Despite involving cognitive processes, System 1 is experiential and uses emotion-related associations and intuitions when making decisions. This research is in line with the practical theory of empiricist philosopher David

the extended elaboration likelihood theory (Cacioppo and Petty 1984; Slater and Rouner 2002) argues that when viewers are transported and emotionally immersed in a narrative, they are more likely to absorb embedded messages⁷⁵. Targeting non-cognitive mediators may be particularly effective with teenagers and young adults⁷⁶. Recent brain development research suggest youth lack regulatory skills to weigh in implications of their decisions, with emotions often triggering risky behaviors (Lansford and Banati 2018; Suleiman and Dahl 2012).

By decomposing treatment effects, mediation analysis tests components or mechanisms that underlie the theory of change of interventions. It helps unpack the “black box” of how an intervention worked (Imai et al 2011), a process often ignored by experimental studies (Deaton 2010a and 2010b). From a policy perspective, mediation analysis is used to separate mechanisms of an intervention that are critical to its success from those that are not, with the goal to further target the former to improve program effectiveness (MacKinnon 1994; Schmiede et al. 2009).

Identifying which mediators help adopt safer sexual practices remains an important gap in the broader HIV literature (Jemmott et al. 2015). On the one hand, a series of mediation studies of school and youth programs in South Africa (Jemmott et al. 2015; O’Leary et al. 2012) and the US (Baumler et al. 2012; Schmiede et al. 2009) find that “traditional” psychosocial constructs such as knowledge, attitudes and behavioral intentions played a limited role in

Hume, who argued that passions are the engine of people’s actions. Without them, people would lack all motivation to act or even to reason, practically or theoretically (A Treatise of Human Nature, 1740).

⁷⁵ Lab research provides empirical evidence for transportation positively mediating behavioral intentions (Hall et al. 2011; Moyer Gusé 2008). These studies suggest that the mechanism behind transportation is the reduction of counter-arguing, the thoughts that may dispute persuasive arguments (Benoit 1987). In other words, audiences absorbed by an entertainment program are less likely to question and rebut program messages.

⁷⁶ Because brain development continues to a lesser extent till the mid-20s, researchers (Sawyer et al. 2018) are calling for adolescence to be expanded by health organizations from 10-19 to 10-24 years of age.

mediating behavioral response. On the other hand, quasi-experimental research in Sub-Saharan Africa shows the direction of HIV putative mediators is more complex than what reasoned-oriented behavior models tend to assume. The altruistic assumption that people that learn they are HIV-positive will reduce their risky sexual behaviors is one example. Gong (2010) and Baird et al. (2014) find HIV-positive people that were adversely surprised by their test results increased, not decreased, risky sexual behaviors.

The current mediation evidence for HIV entertainment-education faces two important limitations. First, this research is often conducted in college labs (in the US), limiting its external validity (Campbell and Stanley 1963). Second, most studies are based on post-intervention surveys that are collected immediately after or within a few days of program exposure, limiting their ability to measure behavior change. Cohort studies consistently find that the “intention-behavior gap” (i.e., the difference between behavioral intentions and actual behaviors) is often substantial, including for HIV testing (Mirkuzie et al. 2011), sexual partner concurrency (McGarritty et al. 2017) and general health behaviors (Sheeran and Webb 2016).

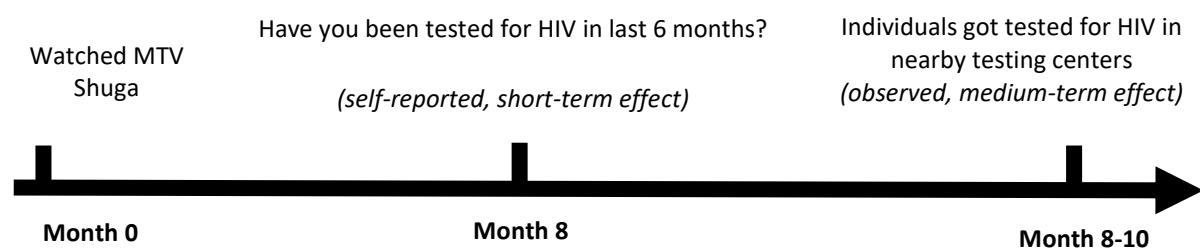
HIV Testing and MTV Shuga

Despite important investments to increase access to HIV testing in Nigeria (National Agency for the Control of AIDS 2015), it is estimated that only 34 percent of HIV-positive people are aware of their HIV-status (UNAIDS Data Book 2017)⁷⁷. Only one in eight sexually-active young adults have been tested in the last 12 months and know their status (UNICEF report, Multiple Indicators Cluster Survey 2016-2017).

⁷⁷ This percent is considerably below the 90-90-90 targets of the UNAIDS Fast Track Strategy (UNAIDS 2014). By 2020, this target intends for 90 percent of people living with HIV to get tested and learn their status; of whom 90 percent are to be on treatment; of whom 90 percent are to be virally suppressed.

Featuring prominent African actors and singers, MTV Shuga is a television drama designed to promote safer sexual practices among African youth. A recent field experiment with 18-25-year-old youth in Nigeria (Banerjee, La Ferrara and Orozco 2018) found that this program had impacts in the intended direction for both HIV psychosocial and behavioral outcomes. Figure 1 displays the timeline of program effects on HIV testing. For a “short-term” testing indicator (collected at the eight-month follow up survey) and a “medium-term” indicator (collected after the follow up survey, between 8 and 10 months after the intervention), the television drama increased this behavioral outcome in both time periods. This mediation analysis aims to understand the cognitive and non-cognitive mediating variables that may be driving program effects in both periods.

Figure 1. Effect timeline on increased HIV testing



This paper asks the following two set of research questions. First, within a causal mediation framework (Hicks and Tingly 2011; Imai et al. 2011), are program effects on HIV testing *separately* mediated by (RQ 1a) knowledge about HIV transmission, (RQ 1b) decrease in stigma and negative attitudes towards HIV-positive people, and (RQ 1c) a higher transportation experience while watching the television program? Second, within an associational framework (Preacher and Hayes 2008), what was the *relative* influence of each mediator on increased HIV testing controlling for the other mediator variables (RQ 2a-2c)?

This confirmatory study⁷⁸ is testing the mentioned theoretical mediators for the following three reasons. First, previous mediation studies of MTV Shuga and other HIV entertainment-education interventions based on observational data suggest that interpersonal communication and knowledge gains (Geary et al. 2007; Keating, Meekers and Adewuyi 2006) and improvements in HIV attitudes (O’Leary et al. 2007) mediated safer sexual behaviors among African youth. Second, the Shuga trial found statistically significant impacts on the three mediating variables, a requirement for mediation analysis (MacKinnon 1994). Third, the three mediators are high-level constructs at the core of the main theories underlying entertainment-education. Excluded mediators, such as identifying with characters, tended to be narrower or to be subsets of the tested constructs, making their comparisons less meaningful (MacKinnon et al. 2002)⁷⁹. It was hypothesized that impacts on HIV testing were positively mediated by each construct. Large indirect effects are rare in psychology because of the prevalence of multiple mediators (Preacher and Kelly 2011).

In Figure 2, the solid lines describe the potential mediating relationships being tested under the mediation framework of indirect and direct effects (MacKinnon et al. 2002)⁸⁰. The study’s theoretical model assumes that program impacts on the three mediating constructs happened immediately after being exposed to the intervention. As a result, these impacts are assumed to precede in time impacts on HIV testing, theoretically fulfilling the time-sequence

⁷⁸ This mediation analysis is *confirmatory* in that it is constructed based on theory, with the goal to empirically test predictions of the theory of change of entertainment-education. In contrast, *exploratory* mediation analysis is used to identify a subset of potential mediators through data-driven strategies (McArdle 2014).

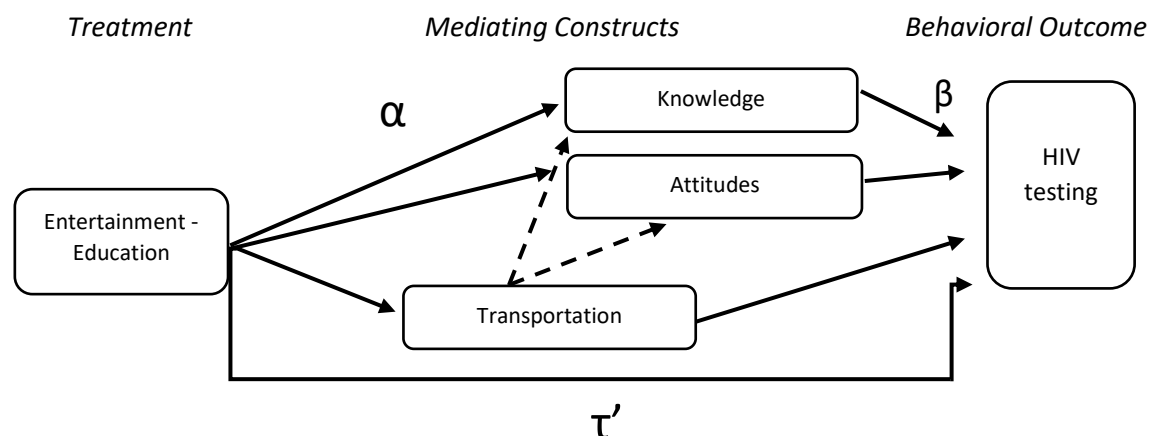
⁷⁹ Although transportation conceptually differs from the other high-level mental constructs, it operationally meets the criteria for a mediator or mediating variable. Namely, it (potentially) transmits the effects it receives from the independent variable to the dependent variable (MacKinnon 2008).

⁸⁰ The indirect effect of a mediator on an outcome is commonly calculated as the product ($\alpha\beta$) of the path between the intervention and the hypothesized mediator (α) and the path between the hypothesized mediator and the outcome (β). Direct effects (τ') measure the influence of other mediators not captured by the tested mediators. The sum of direct and indirect effects equals the total effects on the outcome. The direction of the direct effects can be the same or the opposite of indirect effects.

requirement of mediation frameworks (i.e., mediators subsequently affecting the final outcomes, Baron and Kenny 1986)⁸¹. Multi-stage mediation is hard to interpret in a causal framework as its pathways are not always conceptually clear (e.g., improved knowledge could have directly increased HIV testing or indirectly done so through improved attitudes). For this reason, multi-stage relationships between intermediate variables were not considered. Behavioral intentions were excluded to focus on realized behavior change.

The simplifying assumption of no multi-stage relationships between mediators is likely to underestimate the relative influence of transportation. Lab research suggests increased transportation positively impacts knowledge and attitude outcomes (Hall et al. 2011; Moyer Gusé 2008). This hypothesis is corroborated with the Shuga data, as HIV knowledge and attitudes outcomes were statistically mediated by greater transportation by 14 and 26 percent, respectively (*dotted lines* in Figure 2).

Figure 2. Testing high-level constructs of the theory of change of entertainment-education



⁸¹ This reasonable assumption can be challenged from a stricter point of view for the short-term measure. Impacts on this and the three mediator variables were collected in the same data collection round, the eight-month follow up survey. HIV testing, with its pre- and post-test counseling, is likely to increase HIV knowledge and potentially improve attitudes for individuals that got tested *prior* to the follow-up survey. Because this additional effect on mediators can endogenously inflate mediation quantities, indirect effects for both psychosocial constructs should be read as upper-bound estimates in the short-term model. This potential inflation issue may also affect the transportation variable, though probably to a lesser extent. Transportation items were unrelated to HIV testing and the timing of most transportation questions focused while watching the television drama.

METHODS

Study Design and Sample

The Shuga trial was a field experiment conducted with 18-25-year-old youth in southwest Nigeria. This region has a *generalized* HIV epidemic (UNAIDS 2005 and 2007), with an adult HIV prevalence rate of 2.8 percent (Nigeria Global AIDS Response Report 2017). The intervention was delivered through community film screenings, with no further HIV information provided. To further focus on the effectiveness of the television drama, a placebo rather than a pure control was used to “net-out” the effects of the intervention’s delivery mechanism (i.e., the community screenings). Further details of the Shuga trial can be found in Banerjee, La Ferrara and Orozco (2018) and Orozco (2019).

To give sufficient time for the intervention to potentially impact the HIV testing outcome in the eight-month follow up survey, the sample was restricted to study participants that had already started sexual activity at baseline (i.e., those who hadn’t needed to start a relationship and subsequently get tested for HIV within a short time window. Thus, their inclusion would have likely affected the study’s power). To avoid sample composition bias (Jones and Lang 1980), the sample was further restricted to study participants for whom the trial had data for both HIV testing indicators and the three mediating variables. Ethics approval for the Shuga trial was secured from the National Health Research Ethics Committee of Nigeria and the ethics committees of MIT, Bocconi University, and Oxford University. The Shuga trial was registered at the American Economic Association RCT Registry (<https://www.socialscisceregistry.org/trials/509>).

Mediation Framework

To estimate the indirect causal effect for each hypothesized mediator, three single mediator models were conducted (Hicks and Tingly 2011; Imai et al. 2011). To estimate the relative influence of each controlling for the other mediator variables, one multiple mediator model was conducted (Preacher and Hayes 2008). Running both approaches also helps assess how robust study results are to different methodological approaches (Lu and White 2014).

Single Mediator models

The below equations present the mediation framework for a randomized controlled trial. Although the estimation of treatment impacts on Y_i outcomes (equation 1) and on M_i mediators (equation 2) pose no confounding issues given the random assignment of the treatment (T_i); the estimation of the indirect or mediation effects on final outcomes (equation 3, with M_i corresponding to the multiplicative term $\alpha\beta$ of Figure 2) usually violates the unconfoundedness assumption as e_{i2} and e_{i3} are likely to be correlated (Vansteelandt 2009). When the unconfoundedness assumption does not hold, *causal* estimates will be biased due to omitted variable bias.

1. $Y_i = T_i + e_{i1}$
2. $M_i = T_i + e_{i3}$
3. $Y_i = T_i + M_i + e_{i2}$

By using the potential outcomes framework (Holland 1986), the Hicks and Tingly (2011) method decomposes treatment effects to identify the average causal mediation effect (ACME). This method still doesn't produce a counterfactual outcome for the mediating variable as potential outcomes for controls cannot be observed. To address this limitation,

this method includes a sensitivity test to determine how robust results are to the violations of the sequential ignorability assumption (Rosenbaum 2002)⁸². This sensitivity test shows what correlation between e_{i2} and e_{i3} is needed for the point estimates to be undistinguishable from zero. The test intends to capture how large the effect of an omitted variable needs to be to make a statistically significant effect zero or nonsignificant. Despite methodological improvements, the causal mediation effect should be read as hypothesis-generating research (Bullock, Green and Ha 2010; Gerber and Green 2012)⁸³.

Multiple Mediators model

The multiple mediator analysis was based on structural equations commonly used in mediation studies (Preacher and Hayes 2008); with the indirect effect calculated (MacKinnon et al. 2002) as the product ($\alpha\beta$) of the path between the intervention and the hypothesized mediator (α) and the path between the hypothesized mediator and the outcome (β). See Figure 2. By running the three mediator variables under one model, mediation quantities should theoretically be lower with respect to single mediator models provided mediating coefficients run in the same direction.

⁸² The sequential ignorability assumption applies the unconfoundedness assumption for the treatment assignment followed by the observed mediator, given the observed treatment status and pretreatment confounders. This causal mediation approach assumes that the causal relationship between each mediator variable and the outcome is not confounded by the other mediators post-treatment (e.g., improved knowledge did not affect attitudes). Thus, the sequential ignorability assumption is hard to be met for high-level constructs, and relies on the causal ordering being correct. The latter cannot be verified in the data and can only be supported by theory (Imai et al. 2011).

⁸³ Even with the use of experimental data and assuming high correlation levels between e_{i2} and e_{i3} , the causal mediation framework remains correlational and subject to confounding issues. For example, the extent to which viewers absorb the HIV educational messages or get immersed in the show may be driven by unobservables that may be correlated with HIV testing, biasing mediating estimates (e.g., an individual that is concerned about being infected with HIV may pay greater attention to program messages).

Measures

HIV testing

This study selected HIV testing from a list of behavioral outcomes for three reasons. First, encouraging youth to get tested for HIV is one of the two main messages of MTV Shuga (reducing risky sexual behaviors is the other). Second, items related to HIV testing were included in the HIV knowledge and attitudes measures, improving the empirical link between the psychosocial outcomes and HIV testing. Third, impacts on HIV testing were statistically significant for two different periods, allowing researchers to measure mediating effects in the short- and medium-term. Moreover, the medium-term indicator was independently observed by researchers, overcoming concerns of social desirability bias of self-reported measures (Brener et al. 2003; Minnis et al. 2009).

The *short-term* indicator was collected at the eight-month follow up survey (implementation delays prevented collecting it after six months, the originally planned period). Individuals that reported having gotten tested for HIV at least once before, were asked the last time they got tested. Response options included less than six months ago, 6-12 months ago, 12-13 months ago, and two or more years ago. A binary variable was constructed for individuals that reported getting tested in the last six months. See Figure 1. The *medium-term* measure was collected in the following way. Follow-up respondents from both treatment and control groups were provided with phone-credit incentives (equivalent to US\$7.5) to return two weeks later to the same 80 community centers where the screenings had taken place. Staffed with clinicians from the University College Hospital-Ibadan; these centers now served as “temporary clinics” that collected objective measures for behavior change, including for HIV testing. Clinic attendees were provided with information notes with the address of the closest

HIV testing centers. The information notes included a unique identification number, allowing researchers to observe which individual went to an HIV testing center (not their test results). These notes were collected two months after the follow up survey. It was hypothesized that the treatment group would be more likely to get tested compared to the control group⁸⁴.

Mediator Measures

The eight-month follow-up survey measured program impacts on the three hypothesized mediators. It collected a combined 21 items for the psychosocial variables of HIV knowledge and attitudes. Most of these items were derived from the 2008 Nigeria Demographic and Health survey. This population-level survey has been validated and applied to Nigerian youth since 1990 (Fatusi and Blum 2008; Odu and Foluso 2008). While the knowledge module intended to measure respondents understanding about the correct ways to contract HIV and the benefits of HIV testing; the attitudes module intended to measure stigma and general attitudes towards HIV-positive people (e.g., Would you buy from an HIV-positive shopkeeper?). A few questions related to the main Shuga storylines were included in the survey instrument (e.g., would you allow a boy with HIV to play soccer). A combined 12 items were collected for the transportation measure. Items were adapted from Green and Brock's transportation scale (2000). This module intended to measure the experiential and emotional

⁸⁴ Unfortunately, it was not possible to learn when during this two-month window study participants visited the HIV testing centers (as it was also the case for the short-term indicator, in terms when people got tested between the community screenings and the eight-month follow-up survey). Conversations with clinic staff suggested that most study participants went to get tested within the first couple of weeks. Because the outcome of interest (i.e., subsequently going to an HIV testing center) was not incentivized, and because both experimental groups were incentivized in the same way, the relatively large incentive should not affect the internal validity (Levine and Parkinson 1994) of study findings for this outcome. However, the controlled setting of the temporary clinics and the financial incentives adversely affects the replicability for a normal policy setting and its comparability with the self-reported indicator. While the objective indicator may have induced testing through the provision of information of nearby HIV testing centers (for both treatment and control groups), the self-reported indicator was driven by a passive decision. This inducement may potentially explain the larger effects on the objective measure.

reactions the television program provoked on viewers. Respondents were asked, for example, if they could imagine themselves being part of the story; if the program had affected them emotionally; or if they felt the story had changed their life? Prior to the baseline, a series of pilots were conducted to increase response variation and validate the psychosocial variables in the study population (Presser et al. 2004)⁸⁵.

All three mediating variables were aggregated into indexes for the analysis. The construction of indexes intended (1) to increase statistical power by reducing errors due to random variation through aggregation (Jackson 2003) and (2) to minimize Type-I error inflation due to multiple-hypothesis testing (Young and Karr 2011). These indexes were coded in a way where answers consistent with Shuga's objectives translated into higher values. For their construction, principal component analysis for a family of outcomes was used (Anderson 1963, Jackson 2003). Table A1 presents details of the indicators used in this study, including the number of items, examples of question items and response formats for the index indicators⁸⁶.

Statistical Analysis

To estimate the average causal mediation effect (Hicks and Tingly 2011), three single mediator models were conducted, one for each potential mediating variable. The mediation analysis used probit models for the outcome regressions. Calculation of uncertainty estimates is based

⁸⁵ To the best of the authors' knowledge, the transportation scale has not been validated with Nigerian youth. The validation of the entertainment-education items was not conducted for financial and time constraints, as their validation required conducting screenings.

⁸⁶ Whereas the Cronbach alphas for the indexes of HIV knowledge and transportation were above the commonly accepted rule for internal consistency of 0.70 (Kline 2000; George and Mallery 2003), the Cronbach alpha for the Attitudes index was 0.47. Low alphas suggest that the included items were not closely related. This was partly intentional as psychosocial indexes were intended to be comprehensive measures for the different topics discussed in the three-hour season. These indexes may not be reliable measures for research projects studying more focused interventions and narrower constructs.

on the quasi-Bayesian Monte Carlo approximation (King, Tomz and Wittenberg 2000). Other recent mediation approaches that aim to retrieve causal mediation quantities were considered; however, these were not appropriate for the analysis of binary outcomes (Buis 2011) or lacked sensitivity tests of the sequential ignorability assumption (Buis 2011; Daniel et al. 2011). To estimate mediation effects under the structural equation model (Preacher and Hayes 2008), one multiple mediator model was conducted. Following MacKinnon (et al. 2002), the behavioral outcome was considered to be mediated by the hypothesized mediator if (i) the intervention had a statistically effect on the hypothesized mediator (α), the hypothesized mediator was associated with the behavioral outcome (β) after controlling for the intervention effect, and (iii) the mediated effect was statistically significant ($\alpha\beta$). All point estimates report 95% confidence intervals and were based on two-sided statistical tests.

RESULTS

Randomization Check and Sample Characteristics

Table 1 presents pre-treatment characteristics and outcomes of the study' sample (n=1,894). Just below half were women, with 21 being the average age of respondents. Nine in ten spoke English as their first or second language (Shuga is in English), and nine in ten reported having a sexual partner.

The table shows that treatment and control groups⁸⁷ were balanced at baseline for socio-demographic characteristics, for intermediate outcomes and for the short-term HIV testing indicator (Control=6.1, Treatment=6.9 percent, with the difference not being statistically

⁸⁷ The treatment sample is twice as large as the control sample as the Shuga trial had two treatment arms. While one treatment group only watched MTV Shuga, a second treatment group was exposed to a five-minute videoclip that provided information about prevalence HIV social norms. This mediation analysis combines both treatment arms.

significant $t=0.71$)⁸⁸. Although pre-treatment balance suggest that program impact estimates in the outcome and mediator regressions (equation 1 and 2 respectively) are likely to be unbiased, the study's outcomes of interest (mediated effects in equation 3) primarily depends on the sensitivity tests meeting the unconfoundedness assumption (Rosenbaum 2002).

Table 1. Socio-demographic and HIV outcomes, Pre-intervention characteristics

Variables	Control (n=595)	Treatment (n=1,299)	t-value
Individual characteristics			
Age (years)	21.4 (2.39)	21.4 (2.34)	0.69
Gender (female=1)	42.5%	45.6%	1.27
Speaks English as primary or secondary language	90.8%	90.7%	0.05
Does not have a partner	9.9%	8.6%	0.91
Religion (Muslim=1, Otherwise=0)	40.5%	40.7%	0.09
Hypothetical mediators			
HIV Knowledge index (pca scores)	-0.45 (1.59)	-0.31 (1.67)	1.74
HIV Attitudes index (pca scores)	-0.13 (1.31)	-0.23 (1.30)	0.53
HIV Testing			
Tested for HIV in last 6 months (short-term)	6.1%	6.9%	0.71

For continuous variables, the table presents means and standard deviations. For binary variables, the table shows the sample size and the proportion of the variable in percent terms.

⁸⁸ The objective medium-term measure was only collected at follow up. Given the experimental design of the Shuga trial, the lack of the baseline outcome should not affect the causal interpretation for equation 1 and equation 2 as experimental groups are in expectations identical (Baird et al. 2012; Bruhn and McKenzie 2009; Todd 2014).

Short-term Effects (0 to 8 months after program exposure)

Table 2 presents mediating or indirect effects for the short-term indicator. Panel A presents the average causal mediation effects for the single mediator models (Hicks and Tingly 2011). Under the three models, total effects on increased HIV testing oscillated between 2.6 and 3.1 percentage points (column 1). Direct effects, the influence of *excluded* mediating variables that link the treatment to the outcome impact, represented three-quarters of the total effect. The data suggests that program impacts on HIV testing uptake was greatly mediated by the three mediating variables (column 3). Gains in knowledge about HIV transmission (RQ 1a) mediated 0.6 percentage points (95% CI=0.3 to 1.0 p.p.) or 18.1 percent of the total effect; gains in more favorable attitudes towards HIV-positive people (RQ 1b) mediated 0.6 percentage points (95%CI=0.2 to 1.0 p.p.) or 21.2 percent of the total effect; and greater transportation (RQ 1c) mediated 0.7 percentage points (95%CI=0.2 to 1.3 p.p.) or 26.4 percent of the total effect. Sensitivity results indicate that these causal estimates would only hold for correlation levels between e_{i2} and e_{i3} above 0.10-0.20 (column 5).

Panel B presents the indirect effects for the multiple-mediators model (Preacher and Hayes 2008). It displays the total effect on HIV testing (column 1) and per mediator, its indirect effect size (column 4) and its percent contribution to the total effect when considering other mediator variables (column 5)⁸⁹. The total indirect effect adds up the three mediators (column 6). This associational model provides corroborative evidence that HIV testing was mediated by the three factors. Following MacKinnon et al. (2002), the indirect effect and both paths were statistically significant. Mediation quantities and the percent mediated were slightly

⁸⁹ Indirect or mediated effects are calculated as the product ($\alpha\beta$) of the path between the intervention and the hypothesized mediator (α , column 2) and the path between the hypothesized mediator and the outcome (β , column 3).

lower than estimates obtained through the single mediator models (as expected, as all mediator variables positively mediated treatment effects). The impact of 2.7 percentage points (95%CI=0.00 to 5.3 p.p.) on HIV testing was mediated by gains in knowledge about HIV transmission by 0.4 percentage points (95%CI=0.00 to 0.7 p.p.), by gains in more favorable attitudes towards HIV-positive people by 0.5 percentage points (95%CI=0.1 to 0.8 p.p.), and by gains in transportation by 0.6 percentage points (95%CI=0.00 to 1.1 p.p.). Again, transportation had the strongest mediation influence of the three mediators. While it mediated 22.2 percent of the total effect on HIV testing (RQ 2c), knowledge and attitudes mediated 14.8 and 18.5 percent, respectively (RQ 1-2)⁹⁰. Its greater influence was driven by the intervention's higher impacts on this mediator (path α). The combined indirect effect for the cognitive and non-cognitive mediators was clinically and statistically significant at 1.5 percentage points (95%CI=0.7 to 2.2 p.p.), mediating over half of the effect of MTV Shuga on the short-term indicator (i.e., direct effects mediating less than half).

Despite their large *relative* contributions to the total effect, mediation quantities represent small size effects (Preacher and Kelly 2011) considering the intervention's *absolute* small effect on HIV testing uptake (i.e., 2.6-3.1 percentage points).

⁹⁰ As robustness, the multiple mediators model added an index that measured audiences' identification with characters. The mediating influence of the character identification index was a third of transportation's, with the latter remaining the most influential mediator of this four-mediators model.

Table 2. Mediating effects for short-term outcome

Panel A. Single Mediator Model

Mediator variable	Total effect (95% CI) (1)	Direct effect (95% CI) (2)	Indirect effect (95% CI), ACME (3)	% Mediated (4)	Rho (5)
Knowledge	0.031 (0.004 to 0.056)**	0.025 (-0.002 to 0.051)	0.006 (0.003 to 0.010)**	18.1%	0.1
Attitudes	0.026 (-0.001 to 0.050)	0.020 (-0.006 to 0.045)	0.006 (0.002 to 0.010)**	21.2%	0.2
Transportation	0.026 (-0.000 to 0.051)	0.019 (-0.008 to 0.045)	0.007 (0.002 to 0.013)**	26.4%	0.1

Panel B. Multiple Mediator Model

Mediator variable	Total Effect (95% CI) (1)	α (2)	β (3)	Indirect effect (95% CI), ($\alpha\beta$) (4)	% Mediated (5)	Total Indirect Effects (95% CI) (6)
Knowledge	0.027 (0.000 to 0.053)**	0.340 (0.205 to 0.477)**	0.011 (0.003 to 0.019)**	0.004 (0.000 to 0.007)**	14.8%	0.015 (0.007 to 0.022)**
Attitudes		0.167 (0.044 to 0.291)**	0.026 (0.018 to 0.035)**	0.005 (0.001 to 0.008)**	18.5%	
Transportation		0.934 (0.765 to 1.10)**	0.006 (0.000 to 0.012)**	0.006 (0.000 to 0.011)**	22.2%	

**p<0.0

Medium-term Effects (8 to 10 months after program exposure)

Table 3 presents mediation effects for the medium-term indicator. These results can only provide directional evidence, as *mediation quantities* are no longer statistically significant under either mediation approach (though the *percent they mediate* tends to be statistically significant). Direct effects now mediated over 90 percent of the total effect and became statistically significant.

Total mediation effects represented a fifth of the short-term estimates (Panel B, column 6) and were no longer statistically significant. Interestingly, the relative influence of transportation with respect to both knowledge and attitudes also persisted in the medium-term under the single mediator models (5.4 versus 4.0 and 2.6 percent respectively) and the multiple mediators model (6.3 versus 3.1 and 1.3 percent respectively).

Table 3. Mediating effects for medium-term outcome

Panel A. Single Mediator Model

Mediator variable	Total effect (95% CI) (1)	Direct effect (95% CI) (2)	Indirect effect (95% CI), ACME (3)	% Mediated (4)	Rho (5)
Knowledge	0.031 (0.010 to 0.047)**	0.029 (0.010 to 0.047)**	0.001 (-0.000 to 0.003)	4.0%	0.0
Attitudes	0.032 (0.013 to 0.049)**	0.031 (0.012 to 0.048)**	0.001 (-0.000 to 0.002)	2.6%	0.1
Transportation	0.032 (0.012 to 0.049)**	0.030 (0.010 to 0.048)**	0.002 (-0.002 to 0.006)	5.4%	0.0

Panel B. Multiple Mediator Model

Mediator variable	Total Effect (95% CI) (1)	α (2)	β (3)	Indirect effect (95% CI), ($\alpha\beta$) (4)	% Mediated (5)	Total Indirect Effects (95% CI) (6)
Knowledge	0.032 (0.012 to 0.052)**	0.402 (0.250 to 0.513)**	0.002 (-0.003 to 0.008)	0.001 (-0.001 to 0.003)	3.1%	0.0031 (-0.001 to 0.007)
Attitudes		0.158 (0.021 to 0.294)**	0.003 (-0.004 to 0.010)	0.0004 (-0.00 to 0.002)	1.3%	
Transportation		0.800 (0.613 to 0.987)**	0.002 (-0.003 to 0.007)	0.0017 (-0.002 to 0.01)	6.3%	

**p<0.05

DISCUSSION

To the best of the authors' knowledge, this is the first mediation study of an HIV entertainment-education intervention that measures theorized mediators beyond two weeks after program exposure in a field setting; and the first to complement self-reported outcomes with objective measures for behavior change. Also, this is among the first mediation studies that study time-decay relationships of mediators.

The study provides indicative evidence that HIV testing was strongly mediated by knowledge, attitudes and transportation in the short-term (0 to 8 months), though these mediators played at best a limited role in the medium-term (8 to 10 months). Two methodological approaches gave consistent results across both periods, suggesting that methodological differences are unlikely to qualitatively change study results.

Indirect effects represented over half of short-term effects. Gains in knowledge and attitudes mediated each a fifth of program impacts; with transportation mediating a quarter. Medium-term mediation quantities substantially declined amounting to a fifth of the short-term estimates. The reduced influence of mediators is consistent with findings from a recent meta-analysis of entertainment-education narratives (Orozco, Shen and Cluver 2019), which found directional evidence of a time-decay relationship of entertainment-education narratives on knowledge outcomes⁹¹. Transportation was the most influential mediating variable for the short- and medium-term measures. This finding adds to the growing body of literature on the effectiveness of narratives in the health space. Quasi-experimental studies in the US of

⁹¹ According to that study, the effect size for studies that collected post-intervention data after a month of program exposure (standard mean differences=0.15, 95% CI=0.08-0.22) was approximately a fifth of the effect size of studies that measured effects within a month (SMD=1.14, 95% CI=0.61-1.68), with the difference in effect size being statistically significant.

popular television programs that included messages about cancer and the human papilloma virus (McQueen 2011; Murphy et. 2011; Frank et al. 2015) found that impacts on knowledge, attitudes and behaviors were mediated by increased program transportation.

Taking into account the time-decay influence of knowledge, attitudes and transportation as mediating variables, and that program impacts on HIV testing rates were similar in absolute terms for both periods, the study provides suggestive evidence that tested mediators were “replaced” by others for the medium-term outcome. The mediating influence of untested mediators increased from below half in the short-term to over 90 percent in the medium-term. Identifying these untested mediators is key for designing sexual health interventions to remain effective in the long term (Baumler et al. 2012).

This mediation study has two important strengths. First, the paper helps unpack the black box question of how this theory-based intervention worked, and do so for two different periods (studies that measure time-decay relationships of mediators remain rare). Mediation estimates are subject to fewer confounding issues by using experimental data and the causal mediation framework (Hicks and Tingly, 2011). Moreover, the theoretical framework of the intervention and the timing of the post-treatment data collections help to reasonably rule out the reversal of the causal pathway between mediators and HIV testing (theoretically for the short-term and empirically for the medium-term outcome). Second, this paper provides indicative evidence that entertainment-education mediators are at least as influential in mediating program impacts than traditional psychosocial mediators. Under both mediation approaches and for both measures of HIV testing (for a total of four comparisons), program transportation mediated a larger percent of the total effect with respect to knowledge about HIV transmission and attitudes towards HIV+ people. Moreover, considering that program

transportation mediates impacts on HIV knowledge and attitudes (and that the reversal of the causal pathway is more likely to inflate estimates of both psychosocial mediators), estimates of the relative influence of program transportation are likely to be lower-bounds.

This study had three important limitations. First, study results should be interpreted as associational. The sensitivity results suggest that omitted variable bias is likely to persist in the causal mediation estimates, highlighting post-treatment confounding between the tested high-level constructs (Imai et al. 2011). These theoretical and empirical limitations are reminders that mediation analysis should be interpreted as hypothesis-generating and not as hypothesis-testing research (Gerber and Green 2012; Bullock, Green and Ha 2010). Second, because measurement instruments were not designed having in mind this mediation analysis, instruments may not properly capture the intended constructs. Given that direct and indirect effects run in the same direction, methodological research suggests that this type of measurement error may overestimate mediation estimators (Le Cessie et al. 2012; Fritz, Kenny and MacKinnon 2016). Third, the mediation analysis was not pre-specified in terms of the selected outcome and hypothesized mediators. Despite what may seem to be clear theoretical linkages with established psychological (Ajzen 2014; Bandura 2004) and entertainment-education theories (Cacioppo and Petty 1984; Slater and Rouner 2002), concerns regarding reporting bias cannot be externally validated with a pre-analysis plan (Chan 2005; Gamble et al. 2017).

There are a number of knowledge gaps that follow from this study's findings. Further research is needed to identify effective mediators, especially for long-term behavior change. Improvements in measuring cognitive and non-cognitive mediators should help researchers understand if the theory underlying HIV behavior change campaigns adequately explains the

adoption of sexual behaviors. Future research should study how entertainment-education can complement HIV information-only interventions. For example, in light of the results of Gong (2010) and Baird et al. (2014), which showed that HIV-positive people who were adversely surprised by the results increased their risky sexual behaviors, an important line of research will be to study if complementing current HIV counseling and testing protocols with entertainment-education interventions may promote empathetic emotions and mitigate anti-social behaviors (Murphy et al. 2011; Frank et al. 2015). Similarly, scaling up access to antiretroviral therapy could be complemented with entertainment-education campaigns with the goal of mitigating disinhibition behaviors (de Walque, Kazianga and Over 2012; Kembabazi et al. 2013). To improve their internal and external validity, future mediation studies should test potential “entertainment” mediators through separate experimental arms and do so in field settings (Gerber and Green 2012). This can be accomplished by comparing an entertainment-education versus an HIV information-only treatment arm.

CONCLUSION

Consistent with reasoned-oriented behavior models (Bandura 2004, Ajzen 2014) for HIV testing (Mirkuzie et al. 2011; Kaufman et al. 2015), this study finds that increasing youths’ knowledge about HIV transmission, as well as reducing stigma and negative attitudes towards-people can help promote this behavioral outcome. Findings also suggest that the “entertainment” part of the intervention may have been more important in promoting HIV testing than its “educational” part. This finding is consistent with the theoretical underpinnings of entertainment-education and with modern cognitive science (Bosse, Jonker and Treur 2008; Gutnik et al. 2006; Kahneman 2011) that indicates that emotions play a key role in decision-making. This finding has policy implications for HIV information campaigns in

Sub-Saharan Africa. Considering that HIV awareness is already high among youth in several sub-Saharan countries⁹², the mediating influence of knowledge gains are likely to further diminish. Thus, effective HIV prevention campaigns need to invest in new formats that while they appeal to youth's rational side, should also aim to reach them at a deeper and more emotional level (Singhal et al 2004).

ROLE OF THE FUNDING SOURCE. No external sources supported this study.

ACKNOWLEDGEMENTS. VO wrote the research protocol for the larger Shuga trial. He conducted the mediation analysis and wrote the results. LC contributed to the empirical interpretation of findings and reviewed different versions of the manuscript. VO takes responsibility for the integrity of the data and the accuracy of the data analysis. The findings and conclusions are those of the authors only. The authors note no conflicts of interest.

⁹² According to 2013 Nigeria Demographic Health Survey, among urban 15-24-year-old Nigerian youth, 98 percent had heard about HIV/AIDS and 80 percent knew that condom use prevented HIV infection.

Table A1. HIV Testing and Potential Mediators

Construct	Description	Example items	No. Items	Response format	Alpha
<i>Behavior outcome (Y)</i>					
HIV testing indicator (self-reported)	Behavior outcome for HIV testing collected at follow up. Constructed variable measures testing six months or less prior to the follow-up survey	When was the last time you were tested?	1	Less than 6 months ago, 6-12 months ago, 12-23 months ago, 2 or more years ago	na
HIV testing indicator (observed)	Behavior outcome for HIV testing within two months of follow-up survey	Individual got tested in nearby HIV testing center, tracked through information note provided by research team.	1	Observed	na
<i>Potential Mediators (M)</i>					
HIV Knowledge index	Index capturing individuals' knowledge about HIV transmission and benefits of HIV testing	Number of correct sources of contagion listed, aware of anti-retroviral therapy, knows a second HIV test is necessary after three-month window.	14	True/False, Multiple options, Open questions	0.72
HIV Attitudes towards HIV-positive people index	Index capturing individuals' attitudes toward HIV-positive people	Would not prefer to keep HIV of family member a secret, would reveal HIV status, an HIV-positive boy should play football.	7	Yes/No, 4-points scale (Strongly agree, Agree, Disagree, strongly disagree)	0.47
Transportation index	Index capturing individuals' level of transportation and emotional involvement when watching the program.	You wanted to learn how the story ended, it affected you emotionally, while watching the program, you found your mind wandering.	12	5-point scale (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree)	0.78

CHAPTER 6. SUMMARY AND DISCUSSION OF THESIS FINDINGS

Overview

The objective of this thesis was to study the effectiveness of entertainment-education narratives on promoting safer sexual behaviors among youth. The thesis had three aims: 1) to carry out a meta-analysis of effectiveness trials of the available evidence; 2) to study the effectiveness of the television drama MTV Shuga on youth that engage in risky sexual behaviors in urban Nigeria; and 3) to shed light on the pathways through which this television drama worked in increasing HIV testing. The three papers of this thesis fill several knowledge gaps in the HIV entertainment-education literature, both in developed and developing country contexts, including Nigeria. The current experimental and quasi-experimental evidence on HIV entertainment-education has important methodological limitations.

Most *experimental* studies are limited to college labs in the US (e.g., Moyer-Guse and Nabi 2011), restricting the external validity of their findings (Campbell and Stanley 1963). Most of these studies are reliant on post-intervention surveys collected immediately or a few days after program exposure (e.g., Arroyave 2008; Lapinski 2008). Putting aside likely reporting biases due to Hawthorne effects (McCarney et al. 2007), these studies are restricted to studying effects on intended rather than realized behavior change. Cohort studies consistently find that the “intention-behavior gap”, the difference between intentions and actual behaviors, is substantial, including for HIV testing (Mirkuzie et al. 2011), sexual partner concurrency (McGarrity et al. 2017) and general health behaviors (Sheeran and Webb 2016).

Most *field* studies of HIV entertainment-education interventions are based on observational samples. Their impact estimates are likely to suffer from selection bias (Heckman 1979) as

they often lack control groups (e.g., Jadranin et al. 2015; Pappas-DeLuca 2008; Ramirez et al. 2014) or lack baseline data to carry out difference-in-difference approaches (e.g., Boulay et al. 2002; Collins 2003; Karlyn 2001; Mohammed 2001; O'Leary 2007; Shapiro 2006; Silvestre et al. 2015). Studies based on propensity-score matching strategies often lack sufficient data to control for confounding factors (e.g., Do et al. 2006; Farr 2005).

The first paper of the thesis (Chapter 3) is a meta-analysis that identified ten high-quality field studies in developed and developing countries. It found evidence that indicated that entertainment-education narratives had medium-size effects on increased knowledge and small-size effects on decreased risky sexual behaviors. It found no effects on attitudes, though methodological issues may partly explain the null effects on this outcome (i.e., studies that measured impacts *within* a month of program exposure found negative effects while studies that measured impacts *after* a month found positive effects). This review highlighted several knowledge gaps in the field of entertainment-education research. Firstly, previous research of HIV entertainment-education interventions in Sub-Saharan Africa has studied middle-school students (Dupas 2011; Dupas, Huillery and Seban 2017; Milleliri, Krentel and Rey 2003) and young adults living in rural areas (Vaughan et al. 2000); but no field studies have been carried out for samples representative of the general urban youth population in this region. Secondly, that available evidence tends to include only one gender in their samples (Vaughan et al. 2000 being the exception). Thirdly, no field study of a television drama has studied program effectiveness for high-risk youth in a developing country, crucial for curbing HIV infections in the broader population (Padian et al. 2011; Jones et al. 2014). Finally, previous studies have not collected objective measures for behavior change, including for HIV testing and biomarkers for STIs. The candidate attempted to address these knowledge gaps in the second and third thesis papers.

The second paper of this thesis (Chapter 4) used RCT data of high-risk females (n=1,029), high-risk males (n=1,044) and their combined samples (n=2,073) to study the effectiveness of MTV Shuga on improving HIV knowledge and attitudes and reducing risky sexual behaviors. This paper also measured the moderating effects of religion and media-exposure levels on this high-risk population. These putative moderators have been studied thus far using observational data and mostly in developed countries (e.g., Peterson and Donnerwerth 1998; Collins and Sturgill 2013; Heather et al. 2008). The moderators paper found that the intervention promoted safer sexual behaviors for both genders, with effects being larger for males. It also found no evidence that the intervention was less effective for Muslims or for individuals that watched television on a daily basis.

The third paper of this thesis (Chapter 5) studied the mediating influence on HIV testing of three theoretical mediators: knowledge about HIV transmission, stigma and negative attitudes towards HIV-positive people, and program transportation, or how immersed or engaged a viewer was during and after watching the television drama. Though prior research has studied pathway relationships of HIV school-based interventions in Sub-Saharan Africa (Jemmot et al. 2015; O’Leary et al. 2012) and the US (Baumler et al. 2012; Schmiede et al. 2009), there is no mediation research of HIV entertainment-education interventions in this region. Practically all mediation research of entertainment-education has been conducted in US college labs, and is based on self-reported behaviors. This paper uses an objective measure for HIV testing from a field trial. It found indicative evidence that increases in HIV testing within 8 months of program exposure were greatly mediated by gains in knowledge, attitudes and greater transportation. However, mediation coefficients for the medium-term indicator (8 to 10 months) amounted to a fifth of the effects found in the short-term and were not

statistically significant, suggesting other mediators drove behavioral response in the medium-term.

Papers Two and Three jointly provide evidence about the potential *transportability* of entertainment-education narratives across different populations. While the moderators' paper provides evidence that this communication approach was effective across genders, religious affiliation, and watching-habits; the mediators' paper provides indicative evidence that transportation or program immersion was at least as influential as gains in HIV knowledge and attitudes in mediating program impacts on HIV testing. Engaging audiences at a deeper level, which theoretically should mediate different populations in a similar direction, is at the core of entertainment-education initiatives. Both papers had high external validity (Cook and Campbell 1979), as they evaluated an existing program (generalizability across situations) and their sample was derived from a sample to some extent representative of urban youth in southwest Nigeria (generalizability across people).

This final chapter draws together key findings of the three thesis papers, their strengths and limitations, and discusses policy implications. It concludes with suggestions for future research, dissemination activities and a summary of new research projects the candidate is leading in the field of entertainment-education.

Paper 1: The effectiveness of using entertainment-education narratives to promote safer sexual behaviors of youth: A meta-analysis, 1985-2017

The first paper of this thesis was a meta-analysis that aimed to synthesize global evidence about the effectiveness of entertainment-education narratives in promoting safer sexual

practices among youth. As secondary outcomes, this review measured effect sizes for knowledge and attitudes outcomes.

Description of included studies

The review was able to identify ten studies that met the eligibility criteria following Cochrane Review guidelines. Five studies took place in the US (Vaughan et al. 2000; Milleliri, Krentel and Rey 2003; Dupas 2011; Dupas, Huillery and Seban 2017; Banerjee, La Ferrara and Orozco 2018) and five in Sub-Saharan Africa (Solomon and DeJong 1988; Moyer-Guse and Nabi 2011; Jones, Lacroix and Hoover 2012; Kearney and Levine 2015; Wang and Singhal 2016).

The country's HIV prevalence rate when the studies took place varied between 0.6 and 7.0 percent. The review included complex school-based interventions in middle schools in Cameroon (Dupas, Huillery and Seban 2017), Kenya (Dupas 2011) and Gabon (Milleliri, Krentel and Rey 2003); mobile-based (Jones, Lacroix and Hoover 2012) and information sessions targeted at patients of STI clinics (Solomon and DeJong 1988); television dramas delivered online to college students (Wang and Singhal 2016), and radio and television broadcasts targeted at young adults in rural Tanzania (Vaughan et al. 2000) and teenagers in the US (Kearney and Levine 2015). The review included results of the main paper of the Shuga trial (Banerjee, La Ferrara and Orozco 2018).

The unweighted average age across studies was 19.3 years. Most included studies took place in urban settings. The studied interventions included television screenings (n=5), television broadcasts (n=1), radio broadcasts (n=1), videos delivered through smart phones or personal computers (n=2), and printed comics (n=1). A dose-response relationship was hard to assess due to the limited number of studies and wide dosage variation. While the experimental studies had control over program exposure (as they were conducted in lab settings), the non-

experimental evaluations were mostly field studies and were based on potential exposure to public broadcasts. Controlled exposure was usually under one hour, while potential exposure was above 40 hours. The meta-sample was large, with a total of 23,476 participants (a median of 902 participants per study). The endline outcomes were measured between two weeks and one year after program exposure, with six studies measuring outcomes after three months or more.

Studies were of high quality, with the overall risk-of-bias deemed low using Cochrane Review standards (Higgins and Greene 2011). The most common risk-of-bias was not fully blinding participants and assessors (difficult in the context of behavior change interventions), followed by incomplete outcome data at follow-up. Findings suggest that publication bias was likely to be present, though it should not substantially change the estimates of study effect sizes. For example, Rosenthal's fail-safe test indicates we would need to include 45 null studies (or 5.6 studies null studies per included study) for the combined 2-tailed p-value to exceed 0.05.

Results: Primary Outcomes

This paper found evidence that youth that were exposed to entertainment-education narratives were more likely than comparison participants to engage in safer sexual practices. The study found statistically significant effects for three of the four measured sexual behaviors. Three studies measured impacts on *inter-generational sex*, through the age-gap of sexual partners. The effect size was not significant, however (SMD=0.06, 95% CI=-0.06 to 0.19, four effect sizes). Two studies measured effects on the *number of sexual partners*. The effect size was small and significant (SMD=0.17, 95% CI=0.02 to 0.33, three effect sizes). Six studies measured *unprotected sexual practices*, defined by the review as measures related to condom use and teenage pregnancy. The effect size was small and significant (SMD=0.08, 95% CI=0.03-

0.12, nine effect sizes). Two studies measured *STI testing and management follow up practices*, finding a small and significant effect (SMD=0.29, 95% CI=0.11 to 0.46, two effect sizes).

The meta-analysis included some degree of study heterogeneity as it synthesized data from non-identical interventions and from similar but different outcome indicators. Study heterogeneity was recognized by applying the random-effects model (Hedges and Vevea 1998) and by calculating the I^2 statistics (Higgins et al. 2003; Higgins and Greene 2011). While study heterogeneity was not detected for unprotected sexual practices ($Q = 1.2$, $df(Q)=5$, $I^2=0$), study heterogeneity was high for the other three sexual behaviors: inter-generational sex ($Q = 7.5$, $df(Q)=2$, $I^2=73.3$), number of sexual partners ($Q=1.7$, $df(Q)=1$, $I^2=40.3$), and STI testing and management follow up practices ($Q = 2.4$, $df(Q)=1$, $I^2=57.9$). Despite the large meta-sample, the results of this review should be read cautiously given the small number of included studies and their heterogeneity.

The pooled effect of the four behavior outcomes was small and significant (SMD=0.09, 95% CI=0.06 to 0.12, eighteen effect sizes). Though their pooling may be difficult to conceptually justify, this pooled approach was used for sensitivity tests and to measure time-decay relationships in the subgroup analysis. Sensitivity tests did not reveal single studies exerting influence on the overall effect size or effects changing by removing studies assessed at high risk of bias. The overall effect did not differ when pooling all behavior outcomes using a fixed-effects model. Pooled estimates were approximately half the size found in previous meta-analyses of mass media interventions (Lacroix et al. 2014; Shen & Han 2014). Though a reduction of selection bias may explain this (as this paper excluded studies with weak

identification strategies), a direct comparison is not possible given important differences in terms of included interventions, age of study population, and meta-analyzed outcomes⁹³.

Secondary outcomes

Ten *knowledge* outcomes were retrieved from seven studies. Measurements for most studies were comprehensive, though a few were restricted to measuring specific program messages. A *medium* and significant effect was found for the pooled estimate for knowledge (SMD=0.67, 95% CI=0.32 to 1.02). Four attitudes outcomes were retrieved from four studies. All included studies measured prevention intentions, with one study (Banerjee, La Ferrara and Orozco 2018) also measuring attitudes towards HIV testing and people living with HIV/AIDS.

Meta-analysis subgroup analysis: time decay-relationships

This paper concluded that the available evidence *cannot* provide research insights regarding program moderators, neither through meta-analyzed results or through individual studies. Because factors co-vary across subgroups in meta-analyses, the limited number of included studies prevented conducting the moderators' analysis related to individual characteristics (e.g., gender), study design (e.g., experimental versus quasi-experimental) and intervention characteristics (e.g., television versus radio). Data limitations also prevented individual studies from measuring key moderators, such as the gender or the urban/rural location of audiences. Except for Vaughan et al. (2000), no previous study was powered to study gender heterogeneous effects.

⁹³ To better understand the effectiveness of the storytelling format on youth, the scope of the meta-analysis was narrower with respect to previous reviews of HIV mass media interventions in terms of intervention format (i.e., studies of flyers or public service announcements were excluded) and target population. This review was restricted to 15-24-year old youth, as the psychological development, relationship characteristics and sexual behaviors of this age group is likely to differ from older cohorts (Lansford and Banati 2018; Suleiman and Dahl 2012).

Although the subgroup analysis of the meta-analysis did not find evidence of a time-decay relationship for the pooled behavior estimate (potentially indicating that short-term effects persisted in the medium term), the small number of studies and the lack of studies with individual-level panel data with several post-treatment surveys prevent ruling out alternative explanations. The study found stronger evidence of a time-decay relationship for knowledge outcomes, with the effect size consistently declining with the timing of the follow-up survey. The effect size for studies that collected post-intervention data after a month (SMD=0.15, 95% CI=0.08 to 0.22) was approximately a fifth of the effect size of studies that measured impacts within a month (SMD=1.14, 95% CI=0.61 to 1.68). This difference in effect size was statistically significant ($p=0.001$). This finding is consistent with Brodie et al. (2001), which through repeated cross-sectional data studied knowledge effects of an episode of the television drama “ER” that discussed emergency contraception. That study found evidence of short-term increases in awareness about emergency contraception, though knowledge gains disappeared after two months of the television broadcast.

For attitudes outcomes, the moderators analysis could not reach conclusive findings. In fact, the review provided indicative evidence about the methodological limitations “exit surveys” or of post-treatment surveys collected immediately after program exposure, a common practice in the entertainment- education literature. Negative (opposite) effects were detected for the studies that measured outcomes within a month (SMD=0.20, 95% CI=-0.36 to 0.04) and positive (intended) effects when attitudes were measured after a month (SMD=0.08, 95% CI=0.05 to 0.10). Despite the limited number of studies for each period ($n=2$), this contradictory finding is consistent with findings from individual studies that were based on exit surveys (e.g., Lapinski and Nwulu 2008; Moyer-Gusé and Nabi 2011; Milleliri, Krentel and Rey 2003). Those studies found negative effects on behavioral intentions and attitudes.

This may suggest that self-reported biases are driving this unintended result, as the saliency of the research may make participants perceive it as being “preachy” (Bensley and Wu 1991).

Quality of the Evidence

This meta-analysis concluded that the available evidence for the effectiveness of HIV entertainment-education narratives for knowledge outcomes is strong; with further research needed for both sexual behaviors and attitudes. Using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria (GRADE Working Group 2004), the quality of the evidence was considered moderate for inter-generational sex and unprotected sex, and low for number of sexual partners and STI testing and management. A “moderate” quality rating indicates further research is likely (very likely for a “low” rating) to have an important impact on our confidence in the effect size estimate. For knowledge and attitudes, it was rated high and very low respectively. These ratings may be conservative, as research (Movsisyan, Melendez-Torres, and Montgomery 2016) shows that GRADE guidelines rarely give high ratings to complex and field interventions.

The MTV Shuga trial: Method summary

The first and second papers of this thesis used data from the MTV Shuga trial (n=4,986, effective panel), a cluster randomized control trial led by this doctoral candidate, that took place between 2014 and 2016 in seven cities in southwest Nigeria. This region has an adult HIV prevalence rate of 2.8 percent, below the national average of 3.1 percent (Nigeria Global AIDS Response Report 2015). Thus, the HIV epidemic is classified as *generalized* in the study setting, defined by UNAIDS (2005) as HIV prevalence rates exceeding 1 percent of the general population.

MTV Shuga is an educational television drama broadcasted in all Sub-Saharan countries designed to (i) educate African youth on HIV risks, (ii) reduce stigma against HIV-positive people, and (iii) promote HIV testing uptake and safer sexual behaviors. Shuga messages focus on safe sex rather than abstinence. The trial studied the third season, composed by 8 episodes that lasted a total of three hours.

Study participants were between the ages of 18 and 25. The trial's sample was designed to be as representative of the general urban-youth population: other than age, no exclusion criteria were used. Eighty locations in seven cities were randomly assigned to treatment and control groups. The doctoral candidate is one three Principal Investigators of the broader Shuga trial. This trial was a collaboration amongst the Nigerian government, the Bill and Melinda Gates Foundation, UK Department for International Development, Massachusetts Institute of Technology, Bocconi University and the World Bank research department.

The main paper (Banerjee, La Ferrara and Orozco 2018, see Appendix 2) found significant impacts in the intended direction for knowledge, attitudes and sexual behaviors for the average study participant. Using a series of experimental arms described in the Methods chapter, that paper found that social norms did not mediate outcomes of interest. The outcomes of individuals that were experimentally exposed to a short video clips with social norms information, or where randomly assigned to watch the television program with a friend, did not differ from the outcomes of individuals that were not exposed to either arm. The Shuga trial found experimental evidence that friends of treated viewers had a better understanding about HIV transmission (i.e., information spillovers). However, these friends were not more likely to report improved attitudes and behavior outcomes than the control friends' group.

Paper 2: Does an Educational Television Series Work for the Population Where It Matters Most? A Post-hoc Moderators Analysis of Youth Engaging in High-Risk Sexual Behaviors in Nigeria

This second paper of the thesis aimed to measure the effectiveness of MTV Shuga on high-risk youth (n=2,073). High-risk individuals were defined as those respondents who reported at baseline at least one of the following four HIV-risk factors in the last six months: (i) had unprotected sex in their last sexual encounter; (ii) engaged in sexual concurrency (i.e., simultaneously had two or more sexual partners); (iii) engaged in inter-generational sex, defined as age-disparate relationships with 10 years or more age difference following Balkus et al. (2015) and Schaefer et al. (2017); and (iv) reported symptoms for a sexually transmitted infection. The selection of these high-risk behaviors was informed by considerations relevant to the HIV epidemic in Sub-Saharan Africa. The *subgroup analysis* measured average treatment effects (ATE) for high-risk youth, and ATE per gender. The *effect-modification analysis* intended to understand if impacts *differed* (not ATE) for Christians and for Muslims; or if impacts differed for individuals that watched television on a daily versus at a lower frequency.

Study sample

The high-risk sample was balanced between experimental groups prior to the intervention, including for the self-reported outcomes of HIV testing, condom use and STI symptoms. The moderators paper also found that missing data patterns did not vary systematically by treatment assignment, thus, ruling out attrition bias. While men and women engaged in different risky sexual behaviors at baseline (e.g., men were more likely to report engaging in

sexual concurrency and less likely to get tested for HIV, women were more likely to report engaging in inter-generational sex).

Subgroup analysis of high-risk youth

The moderators paper suggests that MTV Shuga increased the “short-term” indicator for HIV testing (0 to 8 months after program exposure) by 3.3 percentage points (95%CI=0.7 to 5.9), an increase of 32 percent with respect to the control mean. The impact was driven by an increase in HIV testing uptake among females of 4.8 percentage points (95%CI=3.1 to 9.1), which represented a 41 percent increase. The intervention had larger effects on the objective HIV testing indicator collected 8 to 10 months after program exposure. Shuga increased this indicator by 4.8 percentage points (95% CI=1.4 to 8.1), which represented an increase of 184 percent with respect to the control mean of 2.6 percent. This impact was in turn driven by an increase of 5.4 percentage points (95% CI=0.2 to 8.7) among high-risk males, an increase by a factor of five with respect to the control mean of 1.1 percent. The data suggests a reduction of sexual partner concurrency among males. The average number of current sexual partners decreased by -0.14 (95% CI=-0.30 to -0.02), a 10.7 percent reduction. Effectiveness on high-risk males is policy-relevant within the epidemiological context of Nigeria and Sub-Saharan Africa. Compared to female-controls, male-controls were four times less likely to get tested for HIV and six times more likely to report sexual concurrency. Similar gender differences for both behavioral outcomes are observed in population-level studies (Case et al. 2012; Cohen et al. 2011; UNAIDS 2016). Considering their higher potential to contract and spread HIV, behavioral effects on high-risk males may help prevent new HIV infections at the population level (Padian 2011).

The intervention had impacts on both psychosocial outcomes. Statistical increases were observed for the indexes of knowledge about HIV transmission (0.14 standard deviations, 95%CI=0.02 to 0.25) and decreases in stigma and negative attitudes towards HIV-positive people (0.11 standard deviations, 95% CI=0.01 to 0.21). Impacts on females drove psychosocial results. The program increased females' knowledge and attitudes by 0.19 standard deviations (95% CI=0.03 to 0.35) and 0.12 standard deviations (95%CI=0.00 to 0.25), respectively. For males, no statistically significant differences were found for either psychosocial outcome.

The intervention did not impact all outcomes of interest. For the combined and gender sub samples, no effects were found on condom use. These null effects are further confirmed by the lack of effects on behavioral intentions and attitudes towards condom use. Poor access cannot explain these null effects, as four in five respondents noted they could get a condom within ten minutes. This finding contrasts to existing research of HIV entertainment-education interventions that finds increased condom use among patients of STI clinics in the US (Jones et al. 2013) and among student teenager populations in the US (Kearney and Levine 2015) and Sub-Saharan Africa (Dupas 2011; Dupas et al. 2017). These contrasting findings in condom use may potentially be explained by (i) our study participants shifting to safer partners (i.e., by learning the HIV-negative status of partners as well as the observed decrease in sexual concurrency being potentially driven by individuals cutting off the high-risk, non-primary partner) and by (ii) unprotected sex being costlier for individuals in STI treatment (e.g., that knowingly would infect sexual partners) and teenager students (e.g. likely to drop from school due to an unwanted pregnancy). For the Shuga high-risk sample, however, less than 3 percent tested positive for chlamydia and a third currently attended school.

The intervention had no effects on reducing inter-generational sex among females. This result may not be surprising considering that this non-financial intervention was unlikely to have addressed economic factors, which are key determinants on the “supply side” of transactional and inter-generational sex: poorer women are more likely to engage in these high-risk behaviors (Poulin 2007; Dupas et al. 2011; Toska et al. 2015; Schaefer 2017). The lack of effects on inter-generational sex among high-risk females is supported by recent experimental research conducted by the candidate in Malawi. That research finds that by increasing the economic prospects, a vocational training program decreased transactional sex among vulnerable HIV young females. In contrast, the increase of economic prospects increased the “demand” of transactional sex among male trainees (Jones, Orozco and Rascon 2019).

Treatment effects for high-risk youth on psychosocial and behavioral outcomes are similar in magnitude to those reported for the average study participant of the broader Shuga trial (Banerjee, La Ferrara and Orozco 2018), suggesting consistency of program effectiveness between both populations (Alosh et al. 2015). Effect sizes were larger for HIV testing among high-risk youth. While the main paper reported a decrease of chlamydia infections among females, the slightly smaller decrease among high-risk females was not statistically significant. Although the latter may potentially be explained by statistical power limitations due to the use of much smaller samples (Glennerster and Takavarasha 2013), we cannot discard the lack of effects among high-risk females.

Effect-modification results

Results of the effect-modification analysis suggests consistency of program effectiveness for both moderator variables. For both genders, the effect modification analysis suggests no

differential effects by the religion of treated respondents. In other words, despite holding more conservative views than their Christian peers at baseline, MTV Shuga was not less effective in promoting safer sexual practices among Muslim audiences. This quasi-experimental finding provides support for observational research in the US with individuals with strong religious beliefs. That research suggests that mass media can favorably change audiences' attitudes about traditional gender roles and homosexual rights (Peterson and Donnerwerth 1998) and make audiences more accepting of scientific evidence (Collins and Sturgill 2013). However, program effectiveness for other cultural factors may be less clear⁹⁴.

The data also suggests an overall lack of differential effects by television-watching habits. No differential effects were observed for treated male daily viewers with respect to non-daily viewers. For the female sample, the study observes two significant differences though these go in opposite directions. Because coefficients of the interaction terms don't show a general pattern for the two moderators, lack of differential effects and not statistical power limitations (Gerber and Green 2012) may explain potential consistent treatment effects (Alosh et al. 2015; Gerber and Green 2012). This finding contrasts with observational research in the US (Heather et al. 2008) that finds that greater media exposure of positive messages improved health intentions of audiences. One hypothesis that may explain these contrasting results is that because media-consumption levels in our sample are lower compared to those generally observed in the US, our study participants remain under the theorized media-saturation threshold.

⁹⁴ For example, Murphy et al. (2015) finds that an educational narrative with messages about the importance of vaccinating against the human papillomavirus was more effective in promoting vaccination uptake for the population whose culture was portrayed in the program (Mexican-American versus White and African-American).

Paper 3: How did an Entertainment-Education intervention promote HIV testing? A Post-hoc Mediation Analysis of the MTV Shuga Television series

The third paper of this thesis aimed at unpacking the black box to better understand the mechanisms (Deaton 2010a and 2010b; Imai et al 2011) of how MTV Shuga worked in promoting safer sexual behaviors. Specifically, it intended to shed light on mediator pathways that could explain MTV Shuga's impact on increased HIV testing in the short (0-8 months)- and medium-term (8-10 months after program exposure). More specifically, it intended to measure within a causal mediation framework (Hicks and Tingly 2011) the mediating influence of three theoretical mediators central to reasoned-based behavior models (Ajzen 2014; Bandura 2002) and entertainment-education models (Cacioppo and Petty 1984; Slater and Rouner 2002). The study provides indicative evidence that HIV testing was strongly mediated by knowledge, attitudes and transportation in the short-term, though these mediators played a limited role in the medium-term. The two methods gave consistent results across both periods in terms of the relative ranking of the mediators' influence, suggesting that methodological differences are unlikely to qualitatively change study results.

Short-term effects (0 to 8 months after program exposure)

Despite the program's small effects on HIV testing, which oscillated between 2.6 and 3.2 percentage points across the different mediation models, indirect effects represented over half of the short-term effects. The single mediator models suggest that program impacts on HIV testing uptake was mediated by gains in knowledge about HIV transmission (0.6 percentage points, 95% CI=0.3 to 1.0); decreases in stigma and negative attitudes towards HIV-positive people (0.6 percentage points, 95%CI=0.2 to 1.0); and greater transportation during and after watching the television drama (0.7 percentage points, 95%CI=0.2 to 1.3).

However, sensitivity results indicate that causal estimates would only hold for correlation levels between e_{i2} and e_{i3} above 0.10-0.20.

The multiple mediators model (Preacher and Hayes 2008) provides corroborative evidence that HIV testing was mediated by the three mediators. Mediation quantities and the percent mediated were slightly lower than estimates obtained through the single mediator models (as expected, as the multiple mediators model takes into account the other tested mediators). The impact of 2.7 percentage points (95%CI=0.00 to 5.3) on HIV testing was mediated by gains in knowledge (0.4 percentage points, 95%CI=0.00 to 0.7), attitudes (0.5 percentage points, 95%CI=0.1 to 0.8), and transportation (0.6 percentage points, 95%CI=0.00 to 1.1 p.p.). Despite their large *relative* contributions to the total effect, mediation quantities represent small size effects (Preacher and Kelly 2011) considering the intervention's *absolute* small effect on HIV testing uptake (2.6-3.1 percentage points).

Transportation had the strongest mediation influence of the three mediators in the short-term. While it mediated 22.2 percent of the total effect on short-term HIV testing, knowledge and attitudes mediated 14.8 and 18.5 percent, respectively⁹⁵. As noted in the mediators paper, the relative influence of transportation may potentially be higher considering that (i) transportation mediated on average a fifth of program effects on HIV knowledge and attitudes and (ii) short-term mediation estimates for both psychosocial constructs are likely to be inflated for individuals that got tested prior to the follow up survey (i.e., HIV pre- and post-test counselling may improve knowledge and attitude outcomes).

Medium-term effects (8 to 10 months after program exposure)

⁹⁵ As robustness, the multiple mediators model added an index that measured audiences identifying and connecting with characters. The mediating influence of the character identification index was a third of transportation's, with the latter remaining the most influential mediator of this four-mediators model.

Medium-term results can only provide directional evidence, as *mediation quantities* are no longer statistically significant under either mediation approach. Although the *percent contribution* of the three mediator variables remained statistically significant under the single mediator models, their sensitivity tests suggest that the HIV knowledge and transportation indicators are not distinguishable from zero ($Rho=0$). Mediation quantities represented a fifth of the short-term estimates, with the smaller total indirect effect becoming non-significant. Mediation quantities were again consistent between both mediation approaches, with indirect effects being slightly lower under the multiple mediators model. The reduced influence of mediators is consistent across behavioral interventions, including with the findings of the thesis meta-analysis (Orozco, Shen and Cluver 2019), which found supporting evidence of a time-decay relationship of entertainment-education narratives on knowledge outcomes.

Again, transportation remained the most influential mediating variable for the medium-term outcome. The relative influence of transportation with respect to both knowledge and attitudes persisted under the single mediator models (5.4 versus 4.0 and 2.6 percent respectively) and the multiple mediators model (6.3 versus 3.1 and 1.3 percent respectively). This finding is consistent with a series of quasi-experimental studies in the US of popular television programs that included messages about cancer and the human papilloma virus (McQueen 2011; Murphy et. 2011; Frank et al. 2015). These studies found impacts on knowledge, attitudes and behaviors were mediated by program transportation.

Taking into account the time-decay influence of knowledge, attitudes and transportation as mediating variables, and that program impacts on HIV testing rates were similar in absolute terms for both periods, the data suggests tested mediators were “replaced” by others for the

medium-term outcome. The mediating influence of untested mediators, captured by the direct effects in a mediation framework, increased from below half in the short-term to over 90 percent in the medium-term. Identifying these untested mediators is key for designing interventions to remain effective for long-term behavior change.

Strengths of the thesis

The three thesis papers made a series of original contributions to the HIV entertainment-education literature. The first paper was the first meta-analysis conducted for HIV entertainment-education *narratives* (previous reviews of mass media interventions included non-narrative formats such as flyers and public service announcements). Overall, the quality of studies in the meta-analysis was high; and results were robust to a series of methodological decisions. The review excluded college lab studies and focused on effectiveness trials were the type of program exposure followed that of the scaled-up intervention (e.g., “voluntary” take-up for community screenings or “imposed” exposure for school-based interventions), improving the external validity of study findings. The use of study participants from the general population and the use of effectiveness trials mitigate concerns regarding its scalability (Singhal and Rogers 2002).

To the best of the candidate’s knowledge, the second paper was the first to study the effectiveness of an HIV entertainment-education intervention on high-risk individuals using a field experiment complemented with objective measures of behavior change; and the first to study the moderating influence of religion and television-consumption levels using a field experiment in a developing country. This paper was also among the first mediation studies to measure time-decay relationships of mediating variables.

The third paper was the first to study the mediating influence of psychosocial and entertainment-education mediators for an HIV entertainment-education program using a field experiment and objective behavior measures; and the first in a developing country where the post-intervention data was not collected immediately after program exposure, but rather with a time delay.

The moderators and mediators papers shared three important strengths. First, both papers had high external validity (Cook and Campbell 1979), as they evaluated an existing program (generalizability across situations) and their subsamples were derived from a sample that to some extent was representative of urban youth in southwest Nigeria (generalizability across people). However, given the sample's high levels of formal education (11 years) and English fluency (nine in ten spoke it as their first or second language), findings cannot be extrapolated to less educated contexts. Taking into account that the HIV epidemic is generalized in our study setting (above 1 percent of the general population as defined by UNAIDS, 2005), findings may potentially not be applicable to contexts with lower HIV prevalence rates. Second, the use of objective measures for behavior change reduced the risk of social desirability bias in impact estimates, a common concern in the broader HIV literature (Brener et al. 2003; Minnis et al. 2009). Third, selection bias was greatly addressed by the use of experimental data and evidence of pre-treatment balance between treatment and control groups (World Bank 2011).

Each paper also had unique strengths. The moderators paper has two additional strengths. First, it provides strong evidence that MTV Shuga was effective in promoting safer sexual practices among high-risk youth. The internal validity of study findings for this subsample was confirmed by (i) evidence of pre-intervention balance of the baseline data and (ii) adjusted

and unadjusted estimates of program impacts being stable across practically all outcomes of interest (including the objective measures). Second, this paper provides suggestive evidence that MTV Shuga worked across different subpopulations, key for generalizability claims of study findings (Cook and Campbell 1979). This television drama was effective for both high-risk females and high-risk males (though being more effective for males, partly because they began with higher levels of risky sexual behaviors). Program effectiveness did not vary by participants' religion or television-watching habits.

This mediators paper has two additional strengths. First, the paper helps unpack the black box question of how this theory-based intervention worked, and do so for two different periods (studies that measure time-decay relationships of mediators remain rare). Mediation estimates are subject to fewer confounding issues by using experimental data and the causal mediation framework (Hicks and Tingly, 2011). Moreover, the theoretical framework of the intervention and the timing of the post-treatment data collections help to reasonably rule out the reversal of the causal pathway between mediators and HIV testing (theoretically for the short-term and empirically for the medium-term outcome). Second, this paper provides indicative evidence that entertainment-education mediators are as influential in mediating program impacts than traditional psychosocial mediators. Under both mediation approaches and for both measures of HIV testing (for a total of four comparisons), program transportation mediated a larger percent of the total effect with respect to knowledge about HIV transmission and attitudes towards HIV+ people.

Limitations of the thesis

The thesis has a number of limitations. The main limitation of the meta-analysis is the small number of included studies, especially as it prevented studying key moderators. Another

limitation was the inconsistency and the imprecision of effects across studies, which reduced the GRADE quality of evidence score despite the overall high quality of included studies.

The second and third papers have two shared limitations. First, the mediating influence of the “entertainment” component is studied non-experimentally. The lack of an educational “HIV-information-only” control prevents experimentally disentangling the mediating influence of entertainment, including its mediators. For example, the study cannot experimentally test the mediating influence of program transportation or character identification from other potential mediators that were also likely influenced by the program’s messages (e.g., information about the risks of unprotected sex, information about new social norms among youth of treating with respect HIV-positive individuals). Second, the analysis of both papers is not pre-specified. Although both papers provide detailed reasoning behind the selection of study moderators and mediators, and their overall analysis, concerns regarding reporting bias cannot be externally validated with a pre-analysis plan (Chan 2005; Gamble et al. 2017). This limitation is partly addressed by the titles of both papers including the term “Post-hoc Analysis”, and by downgrading their findings to hypothesis-generating and not hypothesis-testing evidence (Spieth et al. 2016).

The second and third thesis papers also have limitations unique to each. The moderators’ study has two additional limitations. First, power limitations are likely affecting detecting statistical findings of this cluster RCT. The subgroup and the effect-modification analyses respectively represent less than half and a quarter of the sample size for which the trial was statistically powered for (i.e., effects by gender for the average study participant). While treatment coefficients were often in the expected direction but not statistically significant (suggesting potential Type II error, effect is true but cannot be statistically detected), the

study cannot rule out the null effect (the null hypothesis is true, and the program did not impact a given outcome). Second, the trial's delivery mechanism (i.e., the community screenings) may have led participants to pay greater attention to the program messages (compared to the attention participants would pay if watching the program on television or in their laptop), confounding this additional attention with the show's impact. This limitation highlights the trade-off faced by mass media research between an "artificial" setting with low external validity but more reliable impact estimates with high internal validity. An individual watching the drama on television or online is more likely to be distracted. For example, a marketing report in urban Nigeria (Consumer Lab Ericsson 2015) finds that half of youth while watching a program online are also likely to be checking their emails, social media or visiting other websites.

The mediators' paper has two additional limitations. First, study results are to be interpreted as associational. The simplifying assumption that mediation effects do not occur between intermediate outcomes may not be a realistic assumption. Despite of the use of experimental data, the sensitivity tests suggest that omitted variable bias may persist in the causal mediation estimates. These theoretical and empirical limitations are reminders that mediation analysis should be interpreted as hypothesis-generating and not as hypothesis-testing research (Gerber and Green 2012; Bullock, Green and Ha 2010)⁹⁶. For example, the extent to which viewers absorb the HIV educational messages or get immersed in the show may be driven by unobservables that are correlated with HIV testing, biasing mediating estimates. For example, an individual that is concerned about being infected with HIV may pay

⁹⁶ Despite these limitations, the study of causal mechanisms or mediators is too important from both a research and policy perspective that it is imperative to continue developing methodological tools than abandoning their study (Imai et al. 2011).

greater attention to program messages. Second, measurement issues may prevent quantifying mediation estimates in a satisfactory manner. Because measurement instruments were not designed having in mind this mediation analysis, instrument items may have not properly captured the intended constructs. Given that direct and indirect effects run in the same direction, methodological research suggests that measurement error is likely to over-estimate mediation estimators (Le Cessie et al. 2012; Fritz, Kenny and MacKinnon 2016).

Directions for future research

There are eight knowledge gaps that follow from this thesis and would benefit from further research. First, more experimental research of edutainment is needed, especially considering the limited effectiveness that systematic reviews find of HIV information campaigns in developing countries (Fonner et al. 2014, McCoy et al. 2010) and Sub-Saharan Africa (Harrison et al. 2010, Michielsen et al. 2010, Mavedzenge et al. 2011, Krishnaratne et al. 2016). Urgent research is needed in Europe, Latin America and South Asia, regions where the meta-analysis found no high-quality studies that met the eligibility criteria of Cochrane Reviews. Future experimental research should be conducted using public television/radio broadcasts or online platforms in both saturated and non-saturated media markets. Research in these more “natural” settings will give a more realistic estimate of program effectiveness as it will incorporate differing distracting factors (and potential intra-household dynamics that emerge from household members together watching and subsequently discussing program content). Moreover, research of media broadcasts and online platforms will allow studying interventions the way they will be scaled up. As a result, this research should provide inputs for estimating cost-benefit and cost-effectiveness measures that policymakers need for their investment decision-making.

Second, improvements in measuring cognitive and non-cognitive mediators should help researchers understand if the theory underlying HIV behavior change campaigns adequately explains decision-making and the adoption of safer sexual behaviors. The thesis finding that program transportation mediated HIV testing may suggest that less traditional mediators may be driving behavioral impacts. Field experiments of interventions aimed at promoting non-cognitive skills should shed further light on their mediating influence with respect to traditional psychosocial constructs (Heckman, Stixrud and Urzua 2006; Jones, Orozco and Rascon 2019).

Third, future research should study how entertainment-education can complement HIV information-only interventions. For example, in light of the results of Gong (2010) and Baird et al. (2014), which showed that HIV-positive people who were adversely surprised by the results increased their risky sexual behaviors, an important line of research will be to study if complementing current HIV counseling and testing protocols with entertainment-education interventions can promote empathetic emotions (Murphy et al. 2011; Frank et al. 2015); indirectly mitigating or reversing the negative effect of learning one's HIV-positive status.

Fourth, research is needed to understand the dose-response and time-decay relationships of entertainment-education, including the minimum amount of exposure to produce policy-relevant effect-sizes in the long-term. It would be interesting to study if a higher dosage is required for impacting outcomes such as condom use or chlamydia infections, outcomes the thesis found limited impacts on. Or, with youth increasingly watching programs online, and episode length rapidly shrinking (Digiday news article 2018), future research should study the effect sizes of shorter but more frequent video clips delivered online. Given that entertainment-education productions tend to be more expensive than traditional HIV

information campaigns, future studies should conduct cost-benefit analyses to understand the minimum audience threshold to justify these larger investments.

Fifth, to provide hypothesis-testing evidence, future studies should (ex-ante) be designed and statistically powered to study program impacts on high-risk youth, including the study of key moderators. For example, to study if an individual's initial levels of HIV knowledge moderate program effectiveness. While the thesis provides indicative evidence for key groups and moderators, hypothesis-testing research is needed for a better informed and sustained scale up. To maximize their epidemiological effectiveness in Sub-Saharan Africa, research is needed in populations at higher risk of contracting and diffusing HIV, including men who have sex with men (Dunkle et al. 2013; Sanders et al. 2013), people who inject drugs (Bowring et al. 2013; Peterson 2013) and sex workers (Baral et al. 2016). To improve the credibility of behavior impacts, future research should complement survey work with objective measurements of behavior change.

Sixth, to improve their internal and external validity, future mediation studies should test potential mediators through separate experimental arms and do so in field settings (Gerber and Green 2012). For example, future research should experimentally study the additional effect that "entertainment" has on top of the "educational" component in mediating safer sexual behaviors. This can be accomplished by comparing an entertainment-education versus an HIV information-only treatment arm. Exploratory mediation analysis (McArdle 2014) is needed to identify the most effective mediator *variables* behind the three tested *indexes*; as well as other commonly studied mediating variables such as self-efficacious beliefs (Bandura 2004) or perception of social norms (Bursztyny, Gonzalez, and Yanagizawa-Drott 2018).

Seventh, important knowledge gaps persist on understanding influential mediators for long-term behavior change. Further research is needed to understand what mediators behind medium-term effects on HIV testing were (i.e., direct effects, the influence of excluded mediating variables that link the treatment to the outcome impact, represented over 90 percent of the medium-term effect). In contrast to one-off or infrequent behaviors, such as HIV testing, the lack of effects on condom use in Paper and Paper 3 suggests research is needed to better understand how to effectively trigger habit-formation for consistent behavior change in the long-term. Population surveys in the US (Shafii et al. 2004) and South Africa (Hendriksen et al. 2007) suggest habit formation is a necessary condition for promoting consistent condom use among youth. Relatedly, the limited mediating influence that social norms had on individuals' attitudes and behaviors, reported by the main paper of the Shuga trial (Banerjee, La Ferrara and Orozco 2018), highlights the need to do further social norms research for both public and private behaviors.

Finally, to test the transportability of the broader concept of entertainment-education, future studies should be carried out across different development sectors; and using different media outlets (e.g., television, radio, mobiles, social media) and content formats (e.g., documentaries, dramas, reality shows, music).

Policy Implications

Results from this thesis suggest four policy recommendations. First, it is hoped that this thesis will convince research donors that mass media campaigns can be rigorously evaluated through creative experimental designs. The evidence base of mass media programs remains dominated by lab experiments with low external validity and observational field studies with low internal validity. In the case of the Shuga trial, the use of placebo screenings helped “net

out” the delivery mechanism and evaluate the effectiveness of Shuga’s content. As development partners are increasingly demanding higher quality evaluations (Banerjee and Duflo 2009; Legovini, Di Maro and Piza 2015), entertainment-education needs to invest in studies with more rigorous identification strategies. More research is needed to ensure effectiveness and sustainability of the next generation of HIV behavior change campaigns.

Second, the small effect sizes found in this thesis should be relevant for policymakers for several reasons. The benefits of preventing even a few HIV infections are substantial considering the human and financial costs of this disease. On the other hand, although program impacts are likely to be small for rare outcomes, their epidemiological importance is not. The program’s impact on increased HIV testing and decreased sexual concurrency among high-risk males (*half* of which reported at baseline engaging in sexual concurrency in the last six months) has the potential to substantially cut new HIV infections at the population level. Moreover, individual small effects can be rapidly aggregated by mass media into meaningful population-level effects. Considering (i) the program’s effectiveness across different segments of urban high-risk youth; and (ii) the economies of scale and the low-costs of television and online distribution, mass media entertainment can potentially be a cost-effective tool for preventing HIV infections at the population level (Wakefield, Loken and Hornik 2010). MTV Shuga is broadcasted in all Sub Saharan countries. It has a strong social media presence, with over 300 million social media interactions in the platforms of Facebook, Instagram, and Twitter. The policy and fiscal advantages of mass media entertainment over more resource-intensive approaches (e.g., cash transfers or face-to-face sensitization interventions) are clearer in countries with generalized HIV epidemics, where UNAIDS (2005 and 2007) recommends targeting all segments of society.

Third, with health behavior models increasingly focusing on macro-social and community-level influences (Zimmerman et al. 2007; Mirkuzie et al. 2011; Kaufman et al. 2015), the thesis provides indicative evidence that individual-level mediators remain playing a key role in health behavior change. On the one hand, and consistent with reasoned-oriented behavior models (Bandura 2004; Ajzen 2014) for HIV testing (Mirkuzie et al. 2011; Kaufman et al. 2015), this study finds that increasing youths' *knowledge* about HIV transmission and about the benefits HIV testing, as well as reducing stigma and decreasing negative *attitudes* towards-people can help promote this behavioral outcome. On the other hand, the mediation findings also suggest that the "entertainment" part of the intervention may have been more important in promoting HIV testing than its "educational" part. For both the short- and medium-term outcomes, the mediating effects of transportation or emotional immersion in the program are at least as influential as the high-constructs of HIV knowledge and attitudes. This finding is consistent with the theoretical underpinnings of entertainment-education (Cacioppo and Petty 1984; Slater and Rouner 2002); and with modern cognitive science (Bosse, Jonker and Treur 2008; Gutnik et al. 2006; Kahneman 2011) and brain development research of youth (Lansford and Banati 2018; Suleiman and Dahl 2012) that indicates that emotions play a key role in decision-making. This finding has policy implications for HIV information campaigns in Sub-Saharan Africa. Considering that HIV awareness is already high among youth in several sub-Saharan countries, the mediating influence of knowledge gains are likely to further diminish. Thus, effective HIV prevention campaigns need to invest in new formats that while appeal to youth's rational side, should also aim to reach them at a deeper and more emotional level.

Fourth, entertainment-education interventions may require complementary and more resource-intensive interventions to be effective with vulnerable populations. Economic and

gender inequalities may potentially explain the null effects on inter-generational sex among high-risk females (i.e., improvements in psychosocial outcomes by this informational intervention cannot be translated into safer sexual practices if economic needs are not addressed). This finding is consistent with observational research in Sub-Saharan Africa that indicates that women are often unable to negotiate consistent condom use with their sexual partners (Chirinda and Peltzer 2014; Hattori 2014). Similarly, research in this region indicates that intergenerational sex is greatly driven by poverty and transactional relationships (Poulin 2007; Dupas et al. 2011; Toska et al. 2015; Schaefer 2017). Complementary interventions aimed at reducing gender power inequalities and economically empowering women may be needed to reduce inter-generational sex (Jones, Orozco and Rascon 2019).

Dissemination

The doctoral candidate has actively disseminated the results of this thesis. Dates and locations are listed in page 3 (Publications and presentations arising from this thesis). He has made presentations to MTV Staying Alive Foundation, the Bill and Melinda Gates Foundation, the UK Department for International Development, the World Bank, and the broader global development community working in entertainment-education and HIV/AIDS. Country stakeholders and policymakers from the Nigerian government were targeted through in-country presentations and workshops organized by the candidate in Abuja (2016) and Lagos (2017). These workshops discussed evaluation results and their implications for future programs and policies in Nigeria. The candidate has made academic presentations at Instituto Tecnológico Autónomo de México-Research and Policy Center (Mexico City, 2016), Johns Hopkins School of Advanced International Studies (Washington DC, 2015 and 2016), Campbell Collaboration “What Works Global” Summit (London, 2016), the International

Communication Association Conference (San Diego, 2017), Middlesex University-Economics Department (London, 2017), and in two seminars at his department (Oxford, 2017 and 2019). The doctoral candidate has made methodological and policy presentations to different units of the World Bank in Washington DC, including the Behavior Change Community of Practice, the Independent Evaluation Group, the Poverty Global Practice, and the candidate's unit, the development economics research group. The candidate wrote a blog in VoxDev (a leading economics blog) and two blogs for the World Bank; and has been interviewed in a series of international development conferences, including the 2017 IMF-World Bank Annual Meetings. As part of his work at the World Bank, the candidate has regularly taught impact evaluation courses to policymakers since 2012.

Next Steps

The doctoral candidate will remain committed to further strengthening the evidence-base of entertainment-education. In great part due to his DPhil dissemination activities, he has been invited to be a technical advisor for mass media research for UNICEF, the Rand Corporation, the Bill and Melinda Gates Foundation, and the Social and Behavior Change Communication Initiative, a multi-donor initiative.

The candidate remains working with the Shuga data on top of his DPhil research. After being rejected by a leading economics journal, the main Shuga paper is being re-drafted to be submitted to another economics journal. In 2019, the candidate co-authored the Shuga paper on domestic violence outcomes. This paper was published in the May 2019 edition of the "Papers and Proceedings of the American Economic Association". In the near future, the candidate will be carrying out an exploratory mediation analysis to identify what mediating variables (including *items* that composed the three tested *indexes*) drove increased HIV

testing in the short- and medium-term. He will also be carrying out a mediation analysis on the “network” sample used in the main paper to estimate program spillovers on friends of main study participants. The main paper found positive spillover effects on knowledge outcomes, with suggestive evidence of negative spillover effects on HIV testing. A hypothesis that may explain the latter result is that friends perceive the “common sexual pool” safer due to observed increases in safer sexual behavior among main study participants. For this new mediation analysis of the network sample, potential mediating variables will include the sexual behaviors of main participants and the perception of HIV-risk among friends.

The DPhil research helped sharpen the research agenda of a new World Bank research program on entertainment-education, launched by the candidate in 2016. Under this program, the candidate is finishing a randomized control trial of radio and printed mass media campaigns aimed at increasing purchases of solar lanterns in rural Senegal (Coville, Orozco, and Reichert 2019). With the primary goal of assessing the transportability of entertainment-education across different development sectors and media outlets, the candidate serves as Principal Investigator for the following three RCTs in Mexico, Nigeria and Brazil.

In Mexico City (n=7,500), a study is testing the effectiveness of an educational *movie* on preventing alcohol and drug addiction among middle-school youth. The study aims to shed light on the role that the delivery mechanism may have on program effectiveness, with half of the schools being exposed to the educational movie in school facilities and the other half in movie theaters (the study partner, Cinopolis, has an active entertainment-education program during weekday mornings). The eight-month follow up survey was collected in the spring of 2019, with preliminary results expected in late 2019.

In northern Nigeria (n=8,200), a study is testing the effectiveness of *community sensitization sessions* designed to improve attitudes towards primary education, especially for girls' education. The intervention includes video clips produced by Discovery Learning Alliance, the foundation of the Discovery Channel. Unlike the Shuga screenings, which were presented as entertainment screenings (e.g., no further HIV information was provided), the sensitization sessions make salient the sanctioning of the social norm message: the community leader opens and closes these sessions with the message that "education is important and aligned with community and religious values". The post-intervention survey is planned for November 2019, with study results expected in 2020.

In Bahia, Brazil (n=6,500), a study is testing the effectiveness of a peer-educator-led program that includes entertainment-education video clips aimed at increasing aspirations and reducing teenage pregnancies of high-school youth. Training is provided to teams of peer educators who then disseminate this information in their schools. The central question of this study is how to find the most effective to "seed" information in a community. The study will experimentally measure the effectiveness of three selection strategies. Peer educators are selected (i) by being nominated by school staff, the status quo selection strategy; (ii) by being the most connected in terms of network centrality measures (Banerjee et al. 2016); and (iii) by being "voted" the most popular student by their school colleagues. The latter two treatment arms were based on a network census collected prior to the baseline. This "seeding" design contrasts to the selection process of the Shuga trial, which studied spillover effects on other community members when inviting an average youth (i.e., youth invited to the community screenings were not selected based on network-centrality measures). The Bahia trial intends to test if the most connected or liked peer educators are more effective in

positively influencing the attitudes and behaviors of their classmates. The post-intervention survey is planned for August-October of 2019, with study results expected in 2020.

For these three new trials, the candidate will be writing moderation and mediation analyses papers. The three trials are powered to study average treatment effects per gender; and their survey instruments have rich modules for a series of mediating outcomes. The candidate hopes that his DPhil research and the new research projects will help him secure an academic position.

It is hoped that this thesis and the above new research will help introduce entertainment-education into mainstream development investments in HIV and other development sectors. The hope is that the moderation analysis may help better target programs; and that findings of the mediation analysis may help better design the content of information campaigns to improve their effectiveness. The goal is that this research will provide rigorous evidence to the argument about the “universal appeal” of engaging storytelling: that a well-told story will appeal and awaken values (e.g., justice) and emotions (e.g., sympathy for those in greatest need) present in the hearts of people regardless of where they live. Consistent with David Hume’s argument that “reasons is the slave of passions”, that these emotions will be the engine for motivating positive behavior change.

REFERENCES

- Abad N, Baack BN, O'Leary A, Mizuno Y, Herbst JH, Lyles CM (2015). A Systematic Review of HIV and STI Behavior Change Interventions for Female Sex Workers in the United States. *AIDS Behav*; 19(9):1701-19. doi: 10.1007/s10461-015-1013-2.
- Ahanonu E and Victor O (2014). Mothers' perceptions of female genital mutilation. *Health Education Research* Pages 1–7
- Airhihenbuwa, C. O. (1999). Of culture and multiverse: Renouncing "the universal truth" in health. *Journal of Health Education*, 30(5), 267-273.
- Ajzen, I. (2014). The theory of planned behaviour is alive and well, and not ready to retire: a commentary on Sniehotka, Penseau, and Araújo-Soares. *Health Psychology Review*.
- Ajzen, I., Albarracín, D., & Hornik, R. (2007). Prediction and change of health behavior: Applying the reasoned action approach. Mahwah, NJ: Lawrence Erlbaum Associates.
- Albarracín D, Johnson BT, Fishbein M, Muellerleile PA (2001). Theories of reasoned action and planned behavior as models of condom use: a meta-analysis. *Psychol Bull*; 127(1):142-61.
- Alosh M, Fritsch K, Huque M, Mahjoob K, Pennello G, Rothmann M, et al. Statistical Considerations on Subgroup Analysis in Clinical Trials. Pages 286-303
- Alter C, Murty S (1997). "Logic modeling: A tool for teaching practice evaluation". *Journal of Social Work Education*. 33 (1): 103–117.
- Anabwani G, Angrist N, Boyda D, Karugaba G, and Suri T (2018). "The Impact of Providing HIV Relative Risk Information on Sexual Behavior of Botswana Youth", JPAL impact evaluation (preliminary findings)
- Anderson, T. W (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics* 34: 122-148.
- Anita C, Martino SC, Collins RL, Elliott MN, Berry SH, Kanouse DE, Miu A (2008). Does Watching Sex on Television Predict Teen Pregnancy? Findings from a National Longitudinal Survey of Youth, *Pediatrics*, 122(5): 1047-1054.
- Asch, Solomon E. (1956), "Studies of independence and conformity: 1. A minority of one against a unanimous majority," *Psychological Monographs* 70(9): 1-70.
- Auvert B., Taljaard D., Lagarde E., Sobngwi-Tambekou J., Sitta R., Puren A (2005). Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med.*;2(11):e298. doi: 10.1371/journal.pmed.0020298.
- Bailey R.C., Moses S., Parker C.B., Agot K., Maclean I., Krieger J.N., Williams C.F., Campbell R.T., Ndinya-Achola J.O (2007). Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *Lancet*. 369(9562):643–656. doi: 10.1016/S0140-6736(07)60312-2.
- Baird S, Garfein R, McIntosh C. Özler B (2012). Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial, *Lancet*, vol. 379: 1320–29.
- Baird S, Gong E, McIntosh C, Özler B (2014). The heterogeneous effects of HIV testing, *Journal of Health Economics*, Vol. 37, issue C, 98-112.
- Baird S, McIntosh C, Özler B. Cash or Condition? (2011). Evidence from a Randomized Cash Transfer

- Program, *Quarterly Journal of Economics*, 126 (4): 1709-1753, 2011.
- Balkus JE, Nair G, Montgomery ET, et al (1999). Age-Disparate Partnerships and Risk of HIV-1 Acquisition among South African Women Participating in the VOICE Trial. *Journal of acquired immune deficiency syndromes*. 2015;70(2):212-217. doi:10.1097/QAI.0000000000000715.
- Bandura A. Health Promotion by Social Cognitive Means, *Health Education & Behavior*, 31 (2): 143–64, 2004.
- Bandura A. Self-Efficacy: Toward a Unifying Theory of Behavioral Change, *Psychological Review* , 84 (2): 191, 1977.
- Bandura, A. (1997). *Self-Efficacy: The Essence of Control*. New York: Freeman Press.
- Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & D. Zillman (Eds.), *Media effects: Advances in theory and research* (pp. 121_153). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bandura, Albert (2004). “Health Promotion by Social Cognitive Means.” *Health Education & Behavior* 31 (2): 143–64.
- Banerjee A and Duflo E (2009). *The Experimental Approach to Development Economics.*, Annual Review of Economics., Vol. 1:151-178.
- Banerjee A, Chandrasekhar A, Duflo E, Jackson M. Gossip: Identifying Central Individuals in a Social Network, MIT Working paper, 2016.
- Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, and Chaudhury S. Hypothesis testing, type I and type II errors. *Ind Psychiatry J*. 2009 Jul-Dec; 18(2): 127–131. doi: 10.4103/0972-6748.62274
- Banerjee A, La Ferrara E, Orozco V (2018). “The Entertaining Way to Behavioral Change: Fighting HIV with MTV”. Draft World Bank Working Paper
- Baral S., Beyrer C., Muessig K., Poteat T., Wirtz A.L., Decker M.R., Sherman S.G., Kerrigan D. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect. Dis*. 2012;12(7):538–549. doi: 10.1016/S1473-3099(12)70066-X.
- Baron RM, Kenny DA. The moderator-mediator distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; 51: 1173–82.
- Baumler E, Glassman J, Tortolero S, et al. Examination of the Relationship between Psychosocial Mediators and Intervention Effects in It’s Your Game: An Effective HIV/STI/Pregnancy Prevention Intervention for Middle School Students. *AIDS Research and Treatment*. 2012;2012:298494. doi:10.1155/2012/298494.
- Benoit W. Argumentation and credibility appeals in persuasion, *Southern Speech Communication Journal*, 52, 2, (181), (1987).
- Bensley, L. S., & Wu, R. (1991). The role of psychological reactance in drinking following alcohol prevention messages. *Journal of Applied Social Psychology*, 21, 1111-1124.
- Berg G and Zia B. Harnessing Emotional Connections to Improve Financial Decisions: Evaluating the Impact of Financial Education in Mainstream Media, *Journal of the European Economic Association*.
- Bertrand, J. T., O’Reilly, K., Denison J., Anhang, R. and Sweat, M. (2006) Systematic review of the effectiveness of mass communication programs to change HIV/AIDS-related behaviors in developing countries. *Health Education Research*, 21(4), pp. 167–197.

- Bosse T, Jonker C, Treur J. Formalisation of Damasio's theory of emotion, feeling and core consciousness. *Consciousness and Cognition*. Volume 17, Issue 1, March 2008, Pages 94-113
- Boulay M, Storey JD, Sood S. Indirect exposure to a family planning mass media campaign in Nepal. *Journal of Health Communications* 2002; Oct-Dec 7(5):379-99.
- Bowring A.L., Luhmann N., Pont S., Debaulieu C., Derozier S., Asouab F., Toufik A., van Gemert C., Dietze P., Stooze M. An urgent need to scale-up injecting drug harm reduction services in Tanzania: prevalence of blood-borne viruses among drug users in Temeke District, Dar-es-Salaam, 2011. *Int. J. Drug Policy*. 2013;24(1):78–81. doi: 10.1016/j.drugpo.2012.08.005.
- Brady, Timothy F., Talia Konkle, George A. Alvarez, and Aude Oliva. "Visual long-term memory has a massive storage capacity for object details." *Proceedings of the National Academy of Sciences* 105, no. 38 (2008): 14325-14329.
- Brener, N., Billy, J., and Grady, W. (2003) "Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence from the Scientific Literature". *Journal of Adolescent Health*, Vol. 33, Num. 6, pp. 436-457.
- Brodie M, Foehr U, Rideout V, Baer N, Miller C, Flournoy R, Altman D. Communicating health information through the entertainment media: A study of the television drama ER lends support to the notion that Americans pick up information while being entertained, *Health Affairs*, 20, 192–199., 2001.
- Brown J. and Keller S. (2000). Can the Mass Media Be Healthy Sex Educators? *Family Planning Perspectives*, 32 (5), 255-257.
- Brown, J. D., and Newcomer, S. F. (1991). Television viewing and adolescents' sexual behavior. *J. Homosexuality* 21: 77–91.
- Bullock J, Green D, Ha S. Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology*. 2010, Vol. 98, No. 4, 550–558
- Bursztyn L, Cantoni D, Funk P, and Yuchtman N. "Polls, the Press, and Political Participation: The Effects of Anticipated Election Closeness on Voter Turnout" 2016.
- Bursztyn, L, Georgy E and Fiorin S. From Extreme to Mainstream: How Social Norms Unravel," 2017, NBER Working Paper No. 23415.
- Bursztyn, L. and R. Jensen (2017). "Social Image and Economic Behavior in the Field: Identifying, Understanding and Shaping Social Pressure". *Annual Review of Economics*, 9, 131.153.
- Bursztyn, L., T. Fujiwara and A. Pallais (2017). "Acting Wife: Marriage Market Incentives and Labor Market Investments". *American Economic Review*, 107(11), 3288-3319.
- Cacioppo J and Petty R (1984) , "The Elaboration Likelihood Model of Persuasion", in *NA - Advances in Consumer Research* Volume 11, eds. Thomas C. Kinnear, Provo, UT: Association for Consumer Research, Pages: 673-675.
- Campbell, D. and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research*. Wadsworth, Chicago.
- Case K, Ghys P, Gouws E, Eaton J, Borquez A, Stover J, Cuchi P, Abu-Raddad L, Garnett GP, Hallett T. *Bulletin of the World Health Organization* (2012); 90:831-838A. doi: 10.2471/BLT.12.102574
- Chan 2005. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005; 330 doi: <https://doi.org/10.1136/bmj.38356.424606.8F>

(Published 31 March 2005).

- Chirinda W and Peltzer K. "Correlates of inconsistent condom use among youth aged 18-24 years in South Africa". *J Child Adolesc Ment Health*. 2014; 26(1):75-82.
- Cho Y, Kalomba D, Mobarak A, and Orozco V. (2013). "Gender Differences in the Effects of Vocational Training in Malawi: Constraints on Women and Drop-Out Behavior". World Bank Working Paper Series, No. 6545
- Cluver L, Meinck F, Boyes M, Orkin M, Sherr L. Structural drivers and social protection: Mechanisms of HIV-risk and HIV-prevention for South African adolescents, *Journal of the International AIDS Society*, vol. 19 2064-6, 2016.
- Cluver, L, Orkin, M, Boyes, M, Gardner, F, & Meinck, F (2011) 'Transactional Sex amongst AIDS-orphaned and AIDS-Affected Adolescents Predicted by Abuse and Extreme Poverty'. *JAIDS* 58 (3) 336-43
- Cluver, L. D., Orkin, F. M., Yakubovich, A. R., & Sherr, L. (2016). Combination Social Protection for Reducing HIV-Risk Behavior Among Adolescents in South Africa. *J Acquir Immune Defic Syndr*. May 1;72(1):96-104. doi: 10.1097/QAI.0000000000000938.
- Coates TJ, Kulich M, Celentano DD, et al. Effect of community-based voluntary counselling and testing on HIV incidence and social and behavioural outcomes (NIMH Project Accept; HPTN 043): a cluster-randomised trial. *The lancet global health*. 2014;2(5):e267-e277. doi:10.1016/S2214-109X(14)70032-4.
- Coates TJ, Richter L, Caceres C. Behavioural strategies to reduce HIV transmission: how to make them work better. *Lancet*. Aug 23 2008;372(9639):669-684.
- Cohen M.S., Chen Y.Q., McCauley M., Gamble T., Hosseinipour M.C., Kumarasamy N., Hakim J.G., Kumwenda J., Grinsztejn B., Pilotto J.H., Godbole S.V., Mehendale S., Chariyalertsak S., Santos B.R., Mayer K.H., Hoffman I.F., Eshleman S.H., Piwowar-Manning E., Wang L., Makhema J., Mills L.A., de Bruyn G., Sanne I., Eron J., Gallant J., Havlir D., Swindells S., Ribaudo H., Elharrar V., Burns D., Taha T.E., Nielsen-Saines K., Celentano D., Essex M., Fleming T.R., HPTN 052 Study Team Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med*. 2011; 365(6):493–505. doi: 10.1056/NEJMoa1105243.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society*, 4, 245_264.
- Collins B. and Sturgill A. The Effects of Media Use on Religious Individuals' Perceptions of Science. *Journal of Media and Religion* 12(4) October 2013
- Collins RL, Elliott MN, Berry SH, Kanouse EE, and Hunter SB. (2003). Entertainment television as a healthy sex educator: The impact of condom efficacy information in an episode of *Friends*. *Pediatrics*, 112, 1115–1121.
- Cook, Thomas D.; Campbell, Donald T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Chicago: Rand McNally College Publishing Company. ISBN 978-0395307908.
- Coville A, Orozco V, and Reichert A. (2019) "Using mass media campaigns to increase adoption of new technologies". World Bank working paper, forthcoming.
- Coville A., Unsch F, Di Maro V, and Zottel S (2015). The impact of Financial Literacy through Feature

- Films: Evidence from a Randomized Controlled Trial in Nigeria, Financial literacy and Education, World Bank conference paper
- Cronbach LJ (1951). "Coefficient alpha and the internal structure of tests". *Psychometrika*. 16 (3): 297–334. doi:10.1007/bf02310555.
- Daniel, R.M., De Stavola, B.L., and Cousens S.N. (2011). *gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula*
- de Walque D, Dow WH, Nathan R, Abdul R, Abilahi F, et al. (2012) Incentivising safe sex: a randomised trial of conditional cash transfers for HIV and sexually transmitted infection prevention in rural Tanzania. *BMJ open* 2: e000747 doi:000710.001136/bmjopen-002011–000747.
- de Walque D, Kazianga H Over M. Antiretroviral Therapy Perceived Efficacy and Risky Sexual Behaviors: Evidence from Mozambique. *Economic Development and Cultural Change*. Vol. 61, No. 1 (October 2012), pp. 97-126 (29 pages)
- Deaton, A. 2010a. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424– 55.
- Deaton, A. 2010b. "Understanding the Mechanisms of Economic Development." *Journal of Economic Perspectives* 24 (3): 3–16
- DellaVigna S and La Ferrara E. Economic and Social Impact of the Media, in *Handbook of Media Economics*. Chapter 19., 2015, pp. Volume 1. Pages 723-768.
- DellaVigna, S. (2007) *Psychology and Economics: Evidence from the Field* (No. w13420), Cambridge, MA: National Bureau of Economic Research.
- DerSimonian R and Laird N. Meta-analysis in clinical trials., *Control Clin Trials*, pp. 7:177–188., 1986.
- Digiday April 27, 2018. <https://digiday.com/media/the-streaming-giants-are-experimenting-with-short-form-but-dont-call-it-a-gold-rush/>
- Dumville J., Torgerson DJ., Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ*. 2006 Apr 22; 332(7547): 969–971. doi: 10.1136/bmj.332.7547.969
- Dunkle K.L., Jewkes R.K., Murdock D.W., Sikweyiya Y., Morrell R. Prevalence of consensual male-male sex and sexual violence, and associations with HIV in South Africa: a population based cross sectional study. *PLoS Med*. 2013;10(6):e1001472. doi: 10.1371/journal.pmed.1001472.
- Dupas P, Huillery E, Seban J. Risk Information, Risk Salience, and Adolescent Sexual Behavior: Experimental Evidence from Cameroon, Stanford University Working Paper January 2017, 2017.
- Dupas P. Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya, *AJ: Applied Economics*, (1), pp.1-36, 2011.
- Dupas P., Orozco V., Robinson J., and Tequame M. The Effects of Cote d'Ivoire's 2011 Civil Conflict on Economic Vulnerability and Transactional Sex. Draft 2016.
- Egger M, Davey Smith G, Schneider M Minder C. Bias in meta-analysis detected by a simple, graphical test , *BMJ*, 315:629-34., 1997.
- Ellis, Paul D. (2010). *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*. United Kingdom: Cambridge University Press.
- El-Sadr WM, Donnell D, Beauchamp G, et al. Financial Incentives for Linkage to Care and Viral Suppression Among HIV-Positive PatientsA Randomized Clinical Trial (HPTN 065). *JAMA Intern*

- Med.* 2017;177(8):1083–1092. doi:10.1001/jamainternmed.2017.2158
- Fatusi A and Blum RW. (2008). Predictors of early sexual initiation among a nationally representative sample of Nigerian adolescents. *BMC Public Health*2008:136. <https://doi.org/10.1186/1471-2458-8-136>
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York: Psychology Press (Taylor & Francis).
- Fishbein, M., & Guinan, M. (1996). Behavioral science and public health: A necessary partnership for HIV prevention. *Public Health Reports*, 3(Suppl. 1), 5-10.
- Fishbein, M., Middlestadt, S. E., Hitchcock, P. J. (1994). Using information to change sexually-transmitted disease-related behaviors: An analysis based on the theory of reasoned action. In R. J. DiClemente & J. L. Peterson (Eds.), *Preventing AIDS: Theories and methods of behavioral interventions*. New York: Plenum.
- Fonner VA, Armstrong KS, Kennedy CE, O'Reilly KR, Sweat MD. School based sex education and HIV prevention in low- and middle-income countries: a systematic review and meta-analysis, *PLoS One*, 2014.
- Frank L, Murphy S, Chatterjee J, Moran M and Baezconde-Garbanati L. (2015) Telling Stories, Saving Lives: Creating Narrative Health Messages, *Health Communication*, 30:2, 154-163, DOI: 10.1080/10410236.2014.974126
- Fritz MS, Kenny DA, MacKinnon DP. The Combined Effects of Measurement Error and Omitting Confounders in the Single-Mediator Model. *Multivariate Behav Res.* 2016;51(5):681-697.
- Geary CW, Burke HM, Castelnuovo L, Neupane S, Sall YB, Wong E, Tucker HT. MTV's Staying Alive global campaign promoted interpersonal communication about HIV and positive beliefs about HIV prevention. *AIDS Education Prevention* 2007 Feb;19(1):51-67.
- Gerber A and Green D. *Field Experiments. Design, Analysis, and Interpretation*. 2012
- Gibbons, F. and Gerrard, M. 1994 *Social-psychological factors in the initiation of adolescent health risk behaviors: a longitudinal study*. Iowa State University
- Glassman JR, Potter SC, Baumler ER, Coyle KK. Estimates of Intraclass Correlation Coefficients From Longitudinal Group-Randomized Trials of Adolescent HIV/STI/Pregnancy Prevention Programs, *Health Education & Behavior*, Vol 42, Issue 4, pp. 545 - 553.
- Glennerster R & Takavarasha K. *Running Randomized Evaluations. A Practical Guide*. Princeton University Press. 2013. ISBN: 9780691159270
- Gong E. Department of Agricultural and Resource Economics, University of California; Berkeley: 2010. HIV testing & risky behavior: The effect of being surprised by your HIV status. Unpublished manuscript.
- GRADE Working Group. *The Grades of Recommendation, Assessment, Development and Evaluation Working Group.* , 2004.
- Graff-Zivin J., Orozco V., Ramirez L., Thirumurthy H. (2015). *Impact Evaluation of Behavior Change Communication Interventions for Kenyan Out-of-school Youth in Nyanza, Kenya. Baseline report*
- Gray R.H., Kigozi G., Serwadda D., Makumbi F., Watya S., Nalugoda F., Kiwanuka N., Moulton L.H., Chaudhary M.A., Chen M.Z., Sewankambo N.K., Wabwire-Mangen F., Bacon M.C., Williams C.F., Opendi P., Reynolds S.J., Laeyendecker O., Quinn T.C., Wawer M.J. Male circumcision for HIV

- prevention in men in Rakai, Uganda: a randomised trial. *Lancet*. 2007;369(9562):657–666. doi: 10.1016/S0140-6736(07)60313-4.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79, 701–721.
- Gunhild, B. Zia, B. (2013). “Harnessing emotional connections to improve financial decisions: evaluating the impact of financial education in mainstream media”. WB Policy Research working paper; no. WPS 6407
- Gutnik LA, Hakimzada AF, Yoskowitz NA, Patel VL. “The role of emotion in decision-making: a cognitive neuroeconomic approach towards understanding sexual risk behavior. *J Biomed Inform*. 2006 Dec;39(6):720-36. Epub 2006 Apr 7.
- Haferkamp CJ. Beliefs about relationships in relation to television viewing, soap opera viewing, and self-monitoring., *Curr. Psychol.* , 18: 193–204., 1999.
- Hall, Alice E., and Cheryl C. Bracken. 2011. “I Really Liked That Movie’: Testing the Relationship between Trait Empathy, Transportation, Perceived Realism, and Movie Enjoyment.” *Journal of Media Psychology: Theories, Methods, and Applications* 23 (2): 90.
- Hall, J.A., Andrzejewski, S.A. & Yopchick, J.E. *J Nonverbal Behav* (2009) 33: 149. <https://doi.org/10.1007/s10919-009-0070-5>
- Harrison A, Newell ML, Imrie J, Hoddinott G. HIV prevention for South African youth: which interventions work? A systematic review of current evidence., *BMC Public Health*, 2010.
- Hartwick J and Barki H (1994). Hypothesis Testing and Hypothesis Generating Research: An Example from the User Participation Literature. *Information Systems Research* 5(4):446-449. <https://doi.org/10.1287/isre.5.4.446>
- Hattori MK. “Trust and condom use among young adults in relationships in Dar es Salaam, Tanzania”. *J Biosoc Sci*. 2014 Sep; 46(5):651-68.
- Heather J, Huang G, Beck V, Murphy S, Valente T. Entertainment-Education in a Media-Saturated Environment: Examining the Impact of Single and Multiple Exposures to Breast Cancer Storylines on Two Popular Medical Dramas. *Journal of Health Communication*, 13:808–823, 2008
- Heckman J, Stixrud J, Urzua S. Th effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. NBER Working paper 12006
- Heckman J. Sample Selection Bias as a Specification Error, *Econometrica*, 47: 153. , 1979.
- Hedges LV and Vevea JL. Fixed and random effects models in metaanalysis., *Psychol Methods*, 3:486–504., 1998.
- Heffron R, McClelland S, Balkus J, Celum C, Cohen C, MugoN, Baeten J et al. Efficacy of oral pre-exposure prophylaxis (PrEP) for HIV among women with abnormal vaginal microbiota: a post-hoc analysis of the randomised, placebo-controlled Partners PrEP Study. *The Lancet HIV*. Volume 4, Issue 10, October 2017, Pages e449-e456
- Herbst JH, Kay LS, Passin WF, Lyles CM, Crepaz N, Marín BV; HIV/AIDS Prevention Research Synthesis (PRS) Team. A systematic review and meta-analysis of behavioral interventions to reduce HIV risk behaviors of Hispanics in the United States and Puerto Rico. *AIDS Behav*. 2007 Jan;11(1):25-47. Epub 2006 Aug 18.
- Higgins J and Greene (editors) S, *Cochrane Handbook for Systematic Reviews of Interventions* Version

- 5.1.0. The Cochrane Collaboration, 2011., [Accessed October 30 2016].
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses., *BMJ*, 327:557–560., 2003.
- Holland, Paul W. (1986). "Statistics and Causal Inference". *J. Amer. Statist. Assoc.* 81 (396): 945–960. doi:10.1080/01621459.1986.10478354. JSTOR 2289064.
- Hope L. Female Genital Mutilation and Female Genital Cutting. *Encyclopedia of Human Rights*, Vol. 2, pp. 200-213, 2009 ; Northeastern University School of Law Research Paper No. 40-2009. Available at SSRN: <http://ssrn.com/abstract=1480323>
- Hume, David (1740). *A Treatise of Human Nature* (1978 ed.). Oxford: Clarendon P. p. 415.
- Hust S, Brown J. & L'Engle K (2008). Boys Will Be Boys and Girls Better Be Prepared: An Analysis of the Rare Sexual Health Messages in Young Adolescents' Media, *Mass Communication and Society*, 11:1, 3-23, DOI: 10.1080/15205430701668139
- Iacobucci, Dawn and Posavac, Steven S. and Kardes, Frank R. and Schneider, Matthew and Popovich, Deidre, The Median Split: Robust, Refined, and Revived (2015). *Journal of Consumer Psychology* 25, 4 (2015) 690-704. Available at SSRN: <https://ssrn.com/abstract=2663369> or <http://dx.doi.org/10.2139/ssrn.2663369>
- Iarossi, Giuseppe. 2006. *The Power of Survey Design: A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/6975>
- Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review* Vol. 105, No. 4 November 2011
- Institute of Education Sciences' (IES) 2008. "What Works Clearinghouse". *Procedures and Standards Handbook Version 3.0*.
- Jackson, J. E. 2003. *A User's Guide to Principal Components*. New York: Wiley.
- Jadranin Z, Dedic G, Vaughan F, Grillo M, Suljagic V. The impact of an educational film on promoting knowledge and attitudes toward HIV in soldiers of the Serbian Armed Forces. *Vojnosanit Pregl.* 2015 Jul;72(7):569-75.
- Jane D. Brown (2002) Mass media influences on sexuality, *The Journal of Sex Research*, 39:1, 42-45, DOI: 10.1080/00224490209552118
- Jemmott JB, Stephens A, O'Leary A, et al. Mediation of Effects of a Theory-Based Behavioral Intervention on Self-Reported Physical Activity in South African Men. *Preventive medicine.* 2015;72:1-7. doi:10.1016/j.ypmed.2014.12.022.
- Johnson BT, Carey MP, Marsh KL, Levin KD, Scott-Sheldon LAJ. Interventions to Reduce Sexual Risk for the Human Immunodeficiency Virus in Adolescents, 1985-2000A Research Synthesis. *Arch Pediatr Adolesc Med.*2003;157(4):381–388. doi:10.1001/archpedi.157.4.381
- Jones A., Cremin I., Abdullah F., Idoko J., Cherutich P., Kilonzo N., Rees H., Hallett T., O'Reilly K., Koechlin F., Schwartlander B., de Zaldondo B., Kim S., Jay J., Huh J., Piot P., Dybul M. Transformation of HIV from pandemic to low endemic levels: a public health approach to combination prevention. *Lancet.* 2014;384(9939):272–279. doi: 10.1016/S0140-6736(13)62230-8.
- Jones R, Lacroix L, Hoover D. Soap opera video episodes streamed to smartphones in a randomized

- controlled trial to reduce HIV sex risk in young urban African American/black women., *Journal of the International AIDS Society*, 15, 157, 2012.
- Jones W and Lang J. Sample Composition Bias and Response Bias in a Mail Survey: A Comparison of Inducement Methods. *Journal of Marketing Research*. Vol. 17, No. 1 (Feb., 1980), pp. 69-76
- Jones M, Orozco V, and Rascon E (2019). "Hard Skills or Soft Talk: Unintended consequences of a vocational training and an inspirational talk on childbearing and sexual behavior in vulnerable youth". Royal Economic Society Annual Conference Paper.
- Kahneman, D. "Thinking, Fast and Slow". 2011. Farrar, Straus and Giroux
- Kaufman M, Cornish F, Zimmerman R, Johnson B. Health Behavior Change Models for HIV Prevention and AIDS Care: Practical Recommendations for a Multi-Level Approach. *J Acquir Immune Defic Syndr*. 2014 Aug 15; 66(Suppl 3): S250–S258.
- Kearney M and Levine P. Media Influences and Social Outcomes: The Effect of MTV's '16 and Pregnant' on Teen Childbearing, *American Economic Review*, pp. vol. 105, no. 12, December 2015 (pp. 3597-3632), 2014.
- Kembabazi A, Bajunirwe F, Hunt PW, Martin JN, Muzoora C, Haberer JE, et al. (2013) Disinhibition in Risky Sexual Behavior in Men, but Not Women, during Four Years of Antiretroviral Therapy in Rural, Southwestern Uganda. *PLoS ONE* 8(7): e69634. <https://doi.org/10.1371/journal.pone.0069634>
- Kharsany, Ayesha B.M. and Quarraisha A. Karim. 2016. "HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities." *Open AIDS Journal* no. 10, 34-48. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4893541/>.
- Kirkwood, Betty and Sterne, Jonathan (2003). *Medical Statistics*. Second Edition. Blackwell Publishing
Libertine
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London: Routledge
- Kohler HP, Thornton RL (2011) Conditional Cash Transfers and HIV/AIDS Prevention: Unconditionally Promising? *The World Bank Economic Review*: doi: 10.1093/wber/lhr1041.
- Kreuter, M. W., Green, M. C., Cappella, J. N., Slater, M. D., Wise, M., Storey, D., & Woolley, S. (2007). Narrative communication in cancer prevention and control: A framework to guide research and application. *Annals of Behavioral Medicine*, 33, 221–235.
- Krishnaratne S, Hensen B, Cordes J, Enstone J, Hargreaves J. Interventions to strengthen the HIV prevention cascade: a systematic review of reviews, *Lancet HIV*, vol. 3: e307–17, 2016.
- Kuhlmann AKS, Kraft JM., Galavotti C, Creek TL, Mooki M, and Ntuny R. (2008). Radio role models for the prevention of mother-to-child transmission of HIV and HIV testing among pregnant women in Botswana. *Health Promotion International*, 23(3), 260-268.
- La Ferrara E. Mass media and social change: can we use television to fight poverty? *Journal of the European Economic Association*, 14: 791–827, 2016.
- La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics*. 4 (4): 1-31.
- LaCroix JM, Snyder LB, Huedo-Medina TB, Johnson BT. Effectiveness of mass media interventions for HIV prevention, 1986–2013: a meta-analysis., *J Acquir Immune Defic Syndr*, (suppl 3): S329–40., 2014.
- Langen T. Gender power imbalance on women's capacity to negotiate self-protection against HIV/AIDS

- in Botswana and South Africa. *African Health Sciences*. Vol 5, No 3 (2005)
- Lansford J. and Banati P. A handbook of adolescent development research and its impact on global policy. Oxford University Press 2018
- Le Cessie S, Debeij J, Rosendaal FR, Cannegieter SC, Vandenbroucke JP. Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology*. 2012;23(4):551–560. [PubMed]
- Legovini A, Di Maro V, Piza C. Impact Evaluation Helps Deliver Development Project. Impact Evaluation helps deliver development projects., World Bank Policy Research Working Paper, 2015.
- Lei JH, Liu LR, Wei Q. Circumcision status and risk of HIV acquisition during heterosexual intercourse for both males and females: a meta-analysis. *PLoS One*, pp. 10: e0125436., 2015.
- Levine, G. and Parkinson, S. (1994). *Experimental Methods in Psychology*. Hillsdale, NJ:Lawrence Erlbaum
- Liu, H., Liu, H., Cai, Y. et al. "Money boys, HIV risks, and the associations between norms and safer sex: a respondent-driven sampling study in Shenzhen, China". *AIDS Behav* (2009) 13: 652. <https://doi.org/10.1007/s10461-008-9475-0>
- Lowry D and Shidler J. Prime-time TV portrayals of sex, safe sex, and AIDS: a longitudinal analysis, *Journalism Quarterly*, 70, 628-637., 1993.
- Lu X and White H. "Checks and robustness tests in applied economics". *Journal of Econometrics*, v. 178, 2014, p. 194-206
- MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83-104.
- Mavedzenge, N., Doyle A., and Ross D. (2011) "HIV prevention in young people in sub-Saharan Africa: a systematic review". *J Adolesc Health*. 2011 Dec;49(6):568-86.
- McArdle JJ. Exploratory data mining using decision trees in the behavioral sciences. In: McArdle JJ, Ritschard G, editors. *Contemporary issues in exploratory data mining*. New York, NY: Routledge; 2014. pp. 3–47.
- McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology*. 2007;7:30. doi:10.1186/1471-2288-7-30.
- McCoy SI, Kangwende RA, Padian NS. Behavior change interventions to prevent HIV infection among women living in low and middle income countries: a systematic review. *AIDS Behavior*, 2010.
- McGarrity LA, Senn TE, Walsh JL, Scott-Sheldon LAJ, Carey KB, Carey MP. Psychological Distress Moderates the Intention-Behavior Association for Sexual Partner Concurrency among Adults. *AIDS and behavior*. 2017;21(6):1567-1571. doi:10.1007/s10461-016-1548-x.
- McKenzie, David, 2011. "Beyond baseline and follow-up: the case for more t in experiments," Policy Research Working Paper Series 5639, The World Bank.
- McQueen, A., Kreuter, M. W., Kalesan, B., & Alcaraz, K. I. (2011). Understanding narrative effects: The

- impact of breast cancer survivor stories on message processing, attitudes, and beliefs among African American women. *Health Psychology*, 30, 674–682.
- Michielsen K, Chersich MF, Luchters S, De Koker P, Van Rossem R, Temmerman M. Effectiveness of HIV prevention for youth in sub-Saharan Africa: systematic review and meta-analysis of randomized and nonrandomized trials. *AIDS*, 2010.
- Milleliri J, Krentel A, Rey J. Sensitisation about condom use in Gabon: evaluation of the impact of a comic book, *Sante.*, 13(4):253-64., 2003.
- Mirkuzie AH, Sisay MM, Moland KM, and Åstrøm AN. “Applying the theory of planned behaviour to explain HIV testing in antenatal settings in Addis Ababa - a cohort study”. *BMC Health Services Research* 2011. 11:196
- Mohammed S. Personal communication networks and the effects of an entertainment- education radio soap opera in Tanzania. *Journal of Health Communication*, 2001; 6(2), 137-154.
- Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche P, Devereaux PJ, Elbourne D, Egger M, Altman D. *BMJ* 2010; CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement., *PLoS Med*, 2009.
- Molina JM, Grund B, Gordin F, Williams I, Schechter M, INSIGHT START study group et al. Which HIV-infected adults with high CD4 T-cell counts benefit most from immediate initiation of antiretroviral therapy? A post-hoc subgroup analysis of the START trial. *The Lancet HIV*. Volume 5, Issue 4, April 2018, Pages e172-e180
- Moonseong H, Namhee K and Myles SF. “Statistical power as a function of Cronbach alpha of instrument questionnaire items”. *BMC Medical Research Methodology* 2015;15:86 <https://doi.org/10.1186/s12874-015-0070-6>
- Moran M, Frank L, Zhao N, Gonzalez C, Thainiyom P, Murphy S, and Ball-Rokeach S. (2016) An Argument for Ecological Research and Intervention in Health Communication, *Journal of Health Communication*, 21:2, 135-138
- Moreau C. The Impact of Television and Radio on Reproductive Behavior and on HIV/AIDS Knowledge and Behavior, DHS Analytical Studies. ICF International, p. No. 24, 2011.
- Movsisyan A, Melendez-Torres GJ, Montgomery P. Outcomes in systematic reviews of complex interventions never reached high GRADE ratings when compared to those of simple interventions, *Journal of Clinical Epidemiology*, 2016.
- Moyer-Gusé E., and Nabi R. (2011). Comparing the effects of entertainment and educational television programming on risky sexual behavior. *Health communication*, 26(5), 416-426.
- Moyer-Gusé, E. (2008), "Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages" *Communication Theory*, Volume 18, Issue 3, pages 407–425
- MTV Shuga website, accessed September 2018, <http://mtvshuga.com>
- Munshi, K. and Myaux, J. (2005), “Social Norms and the Fertility Transition,” *Journal of Development*
- Murphy ST, Frank LB, Chatterjee JS, Moran MB, Zhao N, Amezola de Herrera P, Baezconde-Garbanati LA. Comparing the Relative Efficacy of Narrative vs Nonnarrative Health Messages in Reducing

- Health Disparities Using a Randomized Trial. *Am J Public Health*. 2015 Oct;105(10):2117-23.
- Murphy, S. T., L. B. Frank, M. B. Moran, and P. Patnoe-Woodley (2011). Involved, Transported, or Emotional? Exploring the Determinants of Change in Knowledge, Attitudes, and Behavior in Entertainment-Education., *Journal of Communication*, 61, 407-31.
- Neumayer, E., & Plümper, T. (2017). *Robustness Tests for Quantitative Research (Methodological Tools in the Social Sciences)*. Cambridge: Cambridge University Press. doi:10.1017/9781108233590
- Ng BE, Butler LM, Horvath T, Rutherford GW. Population-based biomedical sexually transmitted infection control interventions for reducing HIV infection. *Cochrane Database of Systematic Reviews*. 2011;(3):CD001220.
- Nigeria Demographic and Health Surveys (2008 and 2013), ICF/Macro, Macro International
- Nigeria Global AIDS Response Report 2017
(<http://www.unaids.org/en/regionscountries/countries/nigeria>)
- Nigeria Multiple Indicators Cluster Survey 2016-2017
- Odu B and Foluso F (2008). Knowledge of HIV/AIDS and Sexual Behavior among the Youths in South West Nigeria. *International Journal of Tropical Medicine* 3 (4): 79-84
- O'Leary A, Jemmott JB, Jemmott LS, Bellamy S, Ngwane Z, Icard L, Gueits L. Moderation and mediation of an effective HIV risk-reduction intervention for South African adolescents. *Ann Behav Med*. 2012 Oct;44(2):181-91. doi: 10.1007/s12160-012-9375-4.
- Orozco V, Graff-Zivin J, Sakho A. "Social Mobilization and Peer-Mentoring to Encourage Voluntary Counseling and Testing and Post-Testing Behavior: A Randomized Experiment in Senegal". World Bank working paper. 2018. Forthcoming
- Oxford Dictionary of Islam, The. Ed. John L. Esposito. Oxford Islamic Studies Online. 19-Mar-2019).
- Padian NS, McCoy SI, Karim SSA, Hasen N, Kim J, et al. (2011) HIV prevention transformed: the new prevention research agenda. *The Lancet* 378: 269–278. doi: 10.1016/s0140-6736(11)60877-5
- Page, R., Hammermeister, J. J., and Scanlan, A. (2000). Everybody's not doing it: Misperceptions of college students' sexual activity. *Am. J. Health Behav.* 24: 387–394.
- Paluck L and Ball L, "Social norms marketing aimed at gender-based violence: a literature review and critical assessment" Report conducted for the International Rescue Committee, May 2010
- Paluck, Elizabeth Levy, and Donald P. Green, 2009. "Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda". *American Political Science Review*, Vol. 103, No. 4
- Papa, M., Auwal, M. A., & Singhal, A. (1995). Dialectic of control and emancipation in organizing for social change: A multitheoretic study of the Grameen Bank in Bangladesh. *Communication Theory*, 5, 189-223.
- Papa, M.J. (2000). "Entertainment-education and social change: An analysis of parasocial interaction, social learning, collective efficacy, and paradoxical communication," *Journal of Communication* 50(4): 31–55.
- Pappas-DeLuca KA, Kraft JM, Galavotti C, Warner L, Mooki M, Hastings P, Kilmarx PH. Entertainment-education radio serial drama and outcomes related to HIV testing in Botswana, *AIDS Education & Prevention*, 20(6), 486-503, 2008.

- Penman-Aguilar A, Carter M, Snead MC, Kourtis AP. Socioeconomic disadvantage as a social determinant of teen childbearing in the U.S., *Public Health Rep*, vol. 128(suppl 1):5-22., 2013.
- PEPFAR (2016) 'Nigeria Gender Assessment'
- Perez-Truglia, R. (2017). Political Conformity: Event-Study Evidence from the United States. *Review of Economics and Statistics*, forthcoming.
- Petersen L, and Donnerwerth G. Religion and Declining Support for Traditional Beliefs about Gender Roles and Homosexual Rights. *Sociology of Religion*, Volume 59, Issue 4, 1 December 1998, Pages 353–371.
- Petersen Z., Myers B., van Hout M.C., Plüddemann A., Parry C. Availability of HIV prevention and treatment services for people who inject drugs: findings from 21 countries. *Harm Reduct. J.* 2013;10:13. doi: 10.1186/1477-7517-10-13.
- Preacher K and Kelley K. "Effect Size Measures for Mediation Models: Quantitative Strategies for Communicating Indirect Effects". *Psychological Methods* 2011, Vol. 16, No. 2, 93–115
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav ResMethods*, 40(3), 879-891.
- Presser S., Rothgeb J., Couper M., Lessler J, Martin E, Martin J, and Singer E. *Methods for Testing and Evaluating Survey Questionnaires*. 2004. Hoboken, New Jersey: John Wiley and Sons, Inc.
- Ramirez-Valles J, Kuhns L, Manjarrez D. Tal Como Somos/Just As We Are: An Educational Film to Reduce Stigma towards Gay and Bisexual Men, Transgender Individuals & Persons Living with HIV/AIDS. *Journal of Health Communications*. 2014 Apr; 19(4): 478–492.
- Rhodes RE, de Bruijn GJ. "How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework". *Br J Health Psychol*. 2013 May;18(2):296-309. doi: 10.1111/bjhp.12032.
- Rivis, A. & Sheeran, P. "Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis". *Curr Psychol* (2003) 22: 218. <https://doi.org/10.1007/s12144-003-1018-2>
- Rosenthal R. The 'File drawer problem' and tolerance for null results, *Psychol Bull* , 86: 638–641., 1979.
- Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28, 267–286.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
- Sabido, Miguel. 2002. *El tono: Andanzas teóricas, aventuras prácticas, el entretenimiento con beneficio social*. Mexico City: UNAM.
- Sanders E.J., Okuku H.S., Smith A.D., Mwangome M., Wahome E., Fegan G., Peshu N., van der Elst E.M., Price M.A., McClelland R.S., Graham S.M. High HIV-1 incidence, correlates of HIV-1 acquisition, and high viral loads following seroconversion among MSM. *AIDS*. 2013;27(3):437–446. doi: 10.1097/QAD.0b013e32835b0f81.
- Sawilowsky, S (2009). "New effect size rules of thumb". *Journal of Modern Applied Statistical Methods*. 8 (2): 467–474.
- Sawyer S, Azzopardi P, Wickremarathne D, Patton G. The age of adolescence. *The Lancet Child & Adolescent Health*. Volume 2, Issue 3, March 2018, Pages 223-228

- Schaefer R, Gregson S, Eaton JW, et al. Age-disparate relationships and HIV incidence in adolescent girls and young women: evidence from Zimbabwe. *AIDS* (London, England). 2017;31(10):1461-1470. doi:10.1097/QAD.0000000000001506.
- Schmiege SJ1, Broaddus MR, Levin M, Bryan AD. Randomized trial of group interventions to reduce HIV/STD risk and change theoretical mediators among detained adolescents. *J Consult Clin Psychol*. 2009 Feb;77(1):38-50. doi: 10.1037/a0014513.
- Shafii T, Stovel K, Davis R, Holmes K. Is condom use habit forming? Condom use at sexual debut and subsequent condom use. *Sex Transm Dis* 31(6):366-72 · July 2004
- Sheeran, P. and Webb, T.L. orcid.org/0000-0001-9320-0068 (2016) The Intention–Behavior Gap. *Social and Personality Psychology Compass*, 10 (9). pp. 503-518.
- Shen F and Han J. Effectiveness of entertainment education in communicating health information: a systematic review., *Asian Journal of Communication*., Volume 24, Issue 6, 2014.
- Sherry J. Media Saturation and Entertainment-Education. *Communication Theory*, Volume 12, Issue 2, 1 May 2002, Pages 206–224, <https://doi.org/10.1111/j.1468-2885.2002.tb00267.x>
- Singhal A and Rogers EM. A Theoretical Agenda for Entertainment-Education, *Communication Theory*, 12 (2): 117–35., 2002.
- Singhal, A. & Rogers, E.M. (1999). *Entertainment-Education: A communication strategy for social change*. Mahwah, NJ: Lawrence Earl-baum Associates.
- Singhal, A. (2001). *HIV/AIDS and communication for behavior and social change: Program experiences, examples, and the way forward*. Geneva, Switzerland: UNAIDS.
- Singhal, A., Cody, M. J., Rogers, E. M., & Sabido, M. (2004). "Entertainment-education and social change: History, research, and practice". Mahwah, NJ: Lawrence Erlbaum.
- Solomon M and DeJong W. The impact of a clinic-based educational videotape on knowledge and treatment behavior of men with gonorrhea., *Sexually transmitted diseases*, 15(3), 127, 1988.
- Sood, S. (2002). Audience involvement and entertainment-education. *Communication Theory*, 2, 153_172.
- Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T. Randomized controlled trials – a matter of design. *Neuropsychiatr Dis Treat*. 2016; 12: 1341–1349. doi: 10.2147/NDT.S101938.
- Sterne JA, Egger M, Smith GD. Systematic reviews in healthcare: investigating and dealing with publication and other biases in meta-analysis., *BMJ*, 323: 101–105., 2001.
- Streiner (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency, *Journal of Personality Assessment*, 80, 99-103
- Sullivan G. "A Primer on the Validity of Assessment Instruments". *J Grad Med Educ*. 2011 Jun; 3(2): 119–120. doi: 10.4300/JGME-D-11-00075.1
- Sussman, S., Dent, C., Mestel-Rauch, J., Johnson, C., Hansen, W. and Flay, B. Adolescent nonsmokers, triers and regular smokers' estimates of cigarette smoking prevalence: When do overestimations occur and by whom? 1988. *Journal of Applied Social Psychology*, 18, 537-551
- Thornton R. The Demand for, and Impact of, Learning HIV Status , *American Economic Review*, 98(5): 1829-63. , 2008.
- Tingley, Dustin, and Raymond Hicks. (2011). "Causal Mediation Analysis." *Stata Journal*, Vol. 11, No.

- 4, pp. 609-615.
- Toska, E, Cluver, L, Boyes, M, Pantelic, M & Kuo, C (2015) "From 'sugar daddies' to 'sugar babies': Quantitatively testing the pathway between inequitable sexual relationships, condom use, and adolescent pregnancy in South Africa.", *Journal of Family Planning and Reproductive Health* 12: 1 59-
- UNAIDS 2014. <http://www.unaids.org/en/resources/campaigns/World-AIDS-Day-Report-2014>.
- UNAIDS 2018. <http://aidsinfo.unaids.org/>. Accessed July 2019.
- UNAIDS Global AIDS Update 2016 (http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf)
- UNAIDS. Joint United Nations Programme on HIV/AIDS and World Health Organization. Practical Guidelines for Intensifying HIV Prevention; Towards Universal Access. UNAIDS; Geneva, Switzerland. 2007. ISBN 978 92 9173 557 0 (NLM classification: WC 5032)
- UNAIDS. Joint United Nations Programme on HIV/AIDS. Geneva: 2005. Intensifying HIV prevention: UNAIDS policy position paper. [cited 2013 Mar 15]. Available from: http://data.unaids.org/publications/irc-pub06/jc1165-intensif_hiv-newstyle_en.pdf.
- UNAIDS. Joint United Nations Programme on HIV/AIDS; The Gap Report ISBN: 978-92-9253-062-4.; 2014.
- UNFPA. United Nations Population Fund. Adolescent Pregnancy: A Review of the Evidence report. New York, 2013.
- UNICEF. For Every Child, End AIDS: Seventh Stocktaking Report, Geneva, 2016.
- VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2(4):457–468.
- Vansteelandt S, Lange C. Causation and causal inference for genetic effects. *Hum Genet*. 2012;131(10):1665–1676.
- Vaughan P, Rogers E, Singhal A, Swalehe R. Entertainment-education and HIV/AIDS prevention: a field experiment in Tanzania., *Journal of Health Communications*, pp. 5 Suppl: 81-100., 2000.
- Vernon R, Ojeda G, Murad R. Incorporating AIDS prevention activities into family planning organization in Colombia. *Stud Fam Plann* 1990; 21: 335–43.
- Vidanapathirana J, Abramson MJ, Forbes A, Fairley C, Mass media interventions for promoting HIV testing. *Cochrane Database of Systematic Reviews*, pp. Issue 3. Art. No.: CD004775., 2005.
- Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. *Lancet*. 2010;376(9748):1261---1271.
- Walsh-Childers K and Brown. JD, Adolescents' acceptance of sex-role stereotypes and television viewing., in *Media, Sex, and the Adolescent*. Creskill, NJ: Hampton Press, 1993, p. 11.
- Wang H and Singhal A (2016). East Los High: Transmedia Edutainment to Promote the Sexual and Reproductive Health of Young Latina/o Americans, *American Journal of Public Health*., Vol. 106, No. 6, pp. 1002-1010.
- Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM (2007). Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med*. Nov 22;357(21):2189-94.
- Ward H, Rönn M. Contribution of sexually transmitted infections to the sexual transmission of HIV

- (2010). *Curr Opin HIV AIDS*. Jul;5(4):305-10. doi: 10.1097/COH.0b013e32833a8844.
- Wasserstein R. & Lazar NA (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, *The American Statistician*, 70:2, 129-133, DOI:10.1080/00031305.2016.1154108
- Weiss M, Bloom H, Brock T. A Conceptual Framework for Studying the Sources of Variation in Program Effects. 2013. MDRC
- Westbrook A, Braver TS. Cognitive effort: A neuroeconomic approach. *Cognitive, affective & behavioral neuroscience*. 2015;15(2):395-415. doi:10.3758/s13415-015-0334-y.
- Westfall, P. and S. Young (1993), *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*, vol. 279, John Wiley & Sons.
- White M, 2016. "BCSTATS: Stata module to analyze back check (field audit) data and compare it to the original survey data," *Statistical Software Components S458173*, Boston College Department of Economics.
- WHO. World Health Organization. HIV Data and Statistics. [Online]. <http://www.who.int/hiv/data/en/>
- WHO. World Health Organization. HIV testing and counselling: guiding principles. Consolidated ARV guidelines, June 2013. http://www.who.int/hiv/pub/guidelines/arv2013/clinical/box5_1/en/
- Wilkinson, Leland (1999). "Statistical methods in psychology journals: Guidelines and explanations". *American Psychologist*. 54 (8): 594–604. doi:10.1037/0003-066X.54.8.594.
- Wilson DB and Lipsey MW. *Practical meta-analysis*. Thousand Oaks, California: Sage, 2001.
- World Bank 2011. *Impact Evaluation in Practice*. Gertler P., Martinez S., Premand P., Rawlings L., and Vermeersch C. ISBN: 978-0-8213-8541-8.
- World Bank, *World Development Report "Mind, Society and Behavior"*, 2015, World Bank, Washington, DC
- World Bank. *World Development Report Digital Dividends*, World Bank, Washington, DC, 2016.
- Young, S. S., Karr, A. (2011). "Deming, data and observational studies" (PDF). *Significance*. 8 (3).
- Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93–98
- Zhao, Xinshu, John G. Lynch Jr., and Qimei Chen (2010), "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis," *Journal of Consumer Research*, 37 (Aug), 197-206.
- Zimmerman RS, Noar SM, Feist-Price S, Dekthar O, Cupp PK, Anderman E, Lock S (2007). Longitudinal test of a multiple domain model of adolescent condom use. *J Sex Res*; 44(4):380-94.