STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2019–2020

Can Access To Fast Internet In Africa Promote HIV Awareness And Prevention? - Evidence from Nigeria

Marie Helen Günther (41555)

Abstract. Can access to the global Internet improve health behaviour and outcomes in Africa? Providing health-related information through the Internet – free of charge, tailored specifically to patients' needs and accessible from anywhere, anytime – has proven to enhance users' mental and physical wellbeing in numerous researched cases. However, little is known about this relation in developing countries, where fast and internationally connected Internet arrived only a decade ago. In how far access to fast Internet can also affect health outcomes, specifically HIV Awareness and Prevention, in Africa is therefore the research question of this study. The arrival of submarine cables at the African coast can be considered a technological exogenous shock which provided access to fast Internet to those areas that were located near the terrestrial backbone network. Matching geo-spatial individual DHS survey data from Nigeria with a terrestrial backbone map, and applying a difference-in-difference model, I compare individuals close to the network to those living more remotely, thus, remaining unconnected. The key results are as follows: First, I find access to fast Internet to increase HIV Protection Awareness for the male, young as well as literate population. This result implies that the Internet indeed poses an effective opportunity to disseminate HIV prevention information targeted to this group specifically. Second, no causal relationship can be identified between fast Internet access and HIV Attitude, Transmission Awareness and Prevention. Thus, in all three cases, the Internet appears insufficient to change deeply rooted perceptions and social stigma.

Keywords: HIV Awareness, HIV Prevention, Internet, Submarine Cables, Nigeria

JEL: I12, I15, O31, R11

Supervisor:	Anders Olofsgård
Date submitted:	18 May 2020
Date examined:	27 May 2020
Discussant:	Carolin Thomas
Examiner:	Kelly Ragan

Acknowledgements

I would like to express my gratitude to my supervisor, Associate Professor Anders Olofsgård, for the clear advises, thought-provoking impulses and timely supervision throughout the past months. Moreover, I would like to thank Sciences Po Paris and Stockholm School of Economics for giving me the opportunity to immerse deeply into both of my passions, development studies and economics over the course of this dual degree. Lastly, I wish to thank everyone in my closest surroundings for constant encouragement, support and companionship.

Table of Contents

List of Tables

List of Figures

Acronyms

AfTerFibre	African Undersea and Terrestrial Fibre Optic Cables Map
AIDS	Acquired immune deficiency syndrome
DHS	Demographic and Health Survey
GDP	Gross domestic product
GPS	The Global Positioning System
HIV	Human immunodeficiency viruses
HRH	Human resources for health
IHCA	Interactive health communication applications
ITU	International Telecommunication Unit
MSM	Men who have sex with men
MTCT	Mother-to-child transmission
NHIS	Nigerian Health Insurance Scheme
OECD	Organization for Economic Co-Operation and Development
STI	Sexually transmitted infection
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

1. Introduction

No other invention has affected individual information exchange and delivery more than the rise of the Internet. Accessible anytime from almost anywhere, it enables users to search for answers to their specific questions and gather knowledge from different sources at their individual pace and tailored to their needs. Its potential in providing access to health information and enabling enhanced health outcomes was already identified in the late 1990s across the developed world. Searching websites for medical information and exchanging experiences with peers through forums, online chats or blogs has empowered millions of patients in self-managing their diseases. Through gaining a more thorough understanding of their diagnoses, symptoms and treatment options, they have learned to better use the health care services available, cope with their disease in everyday life and enhance their physical and mental wellbeing (Kalichman et al., 2011). Moreover, by providing information and medical interventions online, countries' public health care costs can be substantially decreased (Cline & Haynes, 2001).

While fast, reliable and affordable Internet has been accessible in most developed countries for over two decades now, it has only reached Sub-Saharan Africa in the past decade. Thus, the question arises, whether the Internet as a cost-efficient information provider can replicate its success story in the health sector of the developing countries. In this context, attention must be drawn to the severe HIV epidemic: While HIV infection rates differ substantially across the continent, Africa as a whole is profoundly affected containing two-thirds of people suffering from the disease worldwide (WHO, 2019). As knowledge about the disease itself, as well as sexual protection methods and transmission risks remains low and is further impaired by social stigma, prejudice and misconceptions, ignorance remains the greatest risk in relation to raising HIV prevalence rates. Thus, reducing new infections through public education and advertising campaigns is considered one of the most effective strategies in combating the epidemic (Mahajan et al., 2008; Earnshaw & Chaudoir, 2009; Gebremedhin et al., 2017).

Consequently, the main research question of this paper is to determine the effect of fast Internet access on HIV Awareness and Prevention in Africa. However, HIV Awareness can be manifold ranging from knowledge of the disease to a thorough understanding of its threats and transmission routes. To incorporate this diversity of awareness into my study, I investigate three different pillars of awareness: First, I measure Attitude towards people living with HIV to sense cultural prejudices, anxiety or discrimination against infected individuals. Second, I focus on Transmission to determine how well the risks of HIV spread are understood. Third, I seek to identify Protection Awareness, estimating to what extent protection methods are known and linked to HIV prevention. While the Internet is theoretically accessible to anyone, the younger generation as well as men rather than women are considered to constitute the most active Internet users in African countries (GSMA, 2019). Moreover, being literate poses a limitation to effectively using the Internet. I therefore test whether these restricted groups experience a particular benefit from accessing the Internet in terms of HIV outcomes.

While focusing on the country of Nigeria, this paper follows the same approach as Hjort & Poulsen (2019) in considering the arrival of the first submarine cables in Africa between 2009 and 2012 as an exogeneous technological shock to measure the impact of fast Internet access.

Hence, this natural experiment permits the identification and testing of a causal relationship between fast Internet access and HIV Awareness and Prevention by applying a difference-in-difference model. For a specific location within a country to be connected to fast Internet access, two factors must be guaranteed: First, a submarine cable from overseas must arrive at a landing point at the country's coastline and second, the specific location within the country must be located close to the already existing backbone infrastructure. Thus, I define the treatment group based on the geographical distance to the terrestrial backbone: households within a 500m radius around this network are considered to be connected to fast Internet. Using data from the Demographic and Health Survey (DHS), I exploit the possibility to match the geo-coded household-level data of survey respondents with a terrestrial network infrastructure map.

To the best of my knowledge I consider this study to be the first one to investigate the global Internet as a health service provider in itself instead of examining a specific Interned-based health intervention in the context of Africa. Moreover, I contribute to the literature by interlinking Internet use with HIV Awareness and Prevention outcomes whereas previous research in Nigeria mainly focused on determining HIV knowledge in general or examining the hazardous effects online activity could impose on sexual behavior (Adebayo et al., 2006; Onasanya et al., 2008).

In the following, this paper constructs a theoretical framework by reviewing previous literature. It then provides a background on the African Internet infrastructure and the HIV prevalence in Nigeria. Subsequently, it describes the data and the methodology. After presenting the results, it tests the models on their robustness. This paper then provides the discussion of results, followed by a short conclusion.

2. Literature Review

Everywhere in the world patients oftentimes do not understand medical information and lack sufficient disease-related knowledge to entirely grasp more complex contexts in medical consultations. Thus, they often fail to confidently engage in enhanced conversations with medical staff about their own health status. Other times they simply feel too embarrassed to discuss sensitive questions and prefer to remain in ignorance of potential health threats instead. This lack of access to comprehensible and tangible health information poses a risk of leaving millions of health care users behind. Here, the Internet emerges as a universal solution: Accessible from almost anywhere, it enables patients to purposefully search for answers to their specific questions and gather knowledge from different sources at their individual pace without feeling ashamed or uncomfortable. Nevertheless, with the increasing importance of the Internet as a health information provider, a new issue evolves: Is online health information truly equally accessible to everyone?

To gain a deeper understanding of the potential of fast Internet access in accelerating HIV Awareness and Prevention, I first seek to investigate the importance of accessible health information for improved health outcomes on a more general basis. Then, I continue this section by examining different situations and experiments where the Internet, specifically, has served as a driver for health outcomes. With regard to these first two parts, I consider health outcomes universally, referring to fundamental health results containing improved levels of awareness, knowledge or satisfaction, prevention methods, health status, disease or treatment processes, mental or physical wellbeing and side effects - no matter the underlying illness. Thus, I first examine the impact that information and especially the Internet have on all health-related outcomes on a more general level. Eventually, in the third and last sub-section, I explore studies where the Internet has been explicitly examined as a promoter of HIV Awareness and Prevention.

Much research discussed below describes the digital divide as it existed in the early years of Internet access in the Western world. This digital divide may have largely diminished in the developed countries with the vast majority of people now reporting online access through at least one device. On the African continent, however, these studies exactly describe how access to information online may be experienced in many countries today, as the evolvement of a digital information-based society is only taking off now in most regions (GSMA, 2019). Moreover, considering the constant change of online programs and web applications and how people make use of them, studies linked to online interventions must be reviewed with caution keeping in mind the permanent alternation of online health offerings.

2.1 The Effect of Accessible Health Information on Health Outcomes

The health sector is both knowledge-based and information driven. To guarantee effective health care, health providers must ensure that accessible, relevant and affordable information is available to all patients. In fact, the comprehensive delivery of health information is a prerequisite to enable ubiquitous access to health for all (Al-Shorbaji, 2012). The link between access to health information and improved health outcomes has been studied in various contexts. Empirical results uniformly establish the importance of patient information to improve patient's knowledge and decision-making. Interestingly, a full range of different information transmission options exists that have proven effective. Nevertheless, they have one aspect in common: They provide an additional information source to the verbal medical consultation by breaking down complex diagnoses and technical terms into a more comprehensible language, thus making health care provision more accessible in a cognitive manner.

Firstly, McPherson et al. (2001) systematically reviewed randomized controlled trials conducted in the Anglosphere and found that written information material in the form of booklets and packages positively affected cancer patients' outcomes through improved health knowledge, recall, symptom management and health care utilization. However, they also emphasized that patients' information-intake preferences diverge implying that information should be tailored specifically to their needs to enhance relevance and usage. Johnson and Sandford (2005) reached the same conclusion when testing the impact on knowledge and satisfaction when providing written additional to verbal information opposed to only verbal consultation to patients being discharged from the hospital in the US or Canada. Secondly, interactive health communication applications (IHCAs) – computer-based information packages offering disease and treatment information, behavior change advice and social support – showed to positively impact knowledge, clinical outcomes, self-efficacy and continuous behavior changes for patients with chronic diseases in the US (Murray et al., 2005). Thirdly, studies from the Anglosphere that investigate interventions designed to improve information provision and participation in decision-making through audiotaping consultations and information materials report positive impacts on knowledge and satisfaction of cancer patients in general and health illiterate[1] people more specifically (Gaston & Mitchell, 2005; Santo et al., 2005). Especially patients claiming low health literacy perceive standard verbal health consultation as inaccessible and could, thus, benefit from more personalized, comprehensible information materials (Ad Hoc Committee on Health Literacy, 1999). Hence, fourthly, a study where decision aids (i.e. booklets and audiotapes) were translated from medical into plain language reported prostate cancer patients in the US to increasingly engage in discussions with doctors and take up an active role in treatment decision-making (Holmes-Rovner et al., 2005).

The accessibility of information materials further rises in significance when considering the complexity of health care systems for at-risk populations worldwide. Health-disadvantaged groups such as elderly populations, ethnic minorities and socially underprivileged not only face lack of medical knowledge but also economic, geographic and language barriers when trying to access health care services (Nelson, 2002; Healy, 2008). Although the range of available information customized to risk groups increases continuously, people within those target groups often lack knowledge about and access to this information (Sligo & Jameson, 2000). Healy (2008) argues that disseminating the right type of information, such as about low-cost health providers, could increase access to health for marginal groups and, thus, close the health-knowledge gap.

In the past two decades, the Internet has become a key source for health information in the developed world making information access and retrieval easier than ever before (Riegelman & Persily, 2001; Skinner et al., 2003; Tonsaker et al., 2014). Government agencies, health practitioners and organizations are using the Internet as a new channel to provide health information online while individual users share health-related experiences and insights on a peer-to-peer level in online communities (Alpay et al., 2009). By overcoming spatial and temporal barriers while providing easily comprehensible written as well as audiotaped information, the Internet has demonstrated vast potential for closing the health-knowledge gap, making health-related information accessible for everyone (Dutta et al., 2019). Studies from Western countries show that patients primarily go online to find additional health information because they were not satisfied with the quantity or quality of information obtained from their health providers or to gain a deeper, more complete understanding of their physician's diagnosis (McMullan, 2006; Rubenstein, 2012). By reading about others' personal experiences and receiving virtual support from patients who have undergone similar diseases and treatments, Internet users obtain first-hand perspectives oftentimes not communicable through health providers (Hartzler & Pratt, 2011; Neal & McKenzie, 2011). Apart from this, Winzelberg et al. (2003), found that an online support community helped women from California suffering from breast carcinoma to cope with depression, cancer-related trauma and perceived stress levels. Moreover, access to information on the Internet has triggered a paradigm shift in patients' interaction with physicians, allowing them to emerge from their passive role in traditional medical models to a more active, participating position (Peña-Purcell, 2008). Similarly, Fox & Rainie (2002) found that patients using the Internet

[1] People can be considered health illiterate if they face illiteracy or learning difficulties, but also due to language barriers or lack of education (Santo et al., 2005).

to seek illness and treatment-related information engaged more in decision making and practiced more health self-management, thus, resuming more control of their health.

Skinner and Norman (2012, p. 1) describe eHealth literacy as "the ability to seek, find, understand and appraise health information from electronic sources and apply knowledge gained to addressing or solving a health problem." In 2019, however, almost half of the world's population was still not connected to the Internet with great discrepancies between the developed and developing countries, but also between urban and rural areas as well as different social classes within countries (Makri, 2019). Based on a series of interviews conducted in the US, Dickerson et al. (2004) found race and education to determine online health-seeking activities, with socially disadvantaged patient groups using the Internet less. Coming to the same conclusion, Gilmour (2007) argues that although technically available to everyone, usage of the Internet and online health information is still largely determined by socio-economic factors, thus supporting already existing differences in health advantages across the Western world. Similarly, Peña-Purcell (2008) discusses the paradox between free, universally available online health information and the remaining inaccessibility for minority groups in the US population. Specifically, she finds a "digital divide" for Hispanics – although being within reach to health information, they tend to be disconnected from these online services. Older, less educated African Americans also belong to those risk groups, lacking necessary skills to access information online (Alpay et al., 2009). Additionally, a study in Israel found highly eHealth literate (mostly young and educated) participants to benefit more from online information searches with regard to improved health behaviors and use of health insurances as well as patient-physician interactions (Neter & Brainin, 2012).

In contrast, Bundorf et al. (2006) found uninsured US patients with chronic conditions and those with spatial barriers to health care access to be more likely to search for information online. Moreover, in multicultural societies like Israel, where different ethnical groups tend to have varying access to health information, disadvantaged groups reportedly have greater motivation to access medical information via the Internet. Consequently, these groups tend to take on self-action and focus towards accessible measures to overcome their health disadvantages (Mesch et al., 2012). Nevertheless, to make online health information available for all users universally, the quality of Internet access is of paramount importance. Here, four factors are decisive to overcome a digital divide: the need for privacy when searching sensitive information, the non-existence of filtering software on public devices limiting specific online searches, time unlimited online access and functionality in terms of bandwidth and accessibility of devices (Skinner et al., 2003). Moreover, Internet users must be able to distinguish credible information sources from those spreading half-knowledge or misinformation (Eastin, 2001).

2.2 The Internet as a Driver of Health Awareness and Outcomes

Access to online health information and interventions has the potential to increase users' self-management and feeling of empowerment. Firstly, online health information programs are estimated to induce individuals to become more active and responsible in searching for medical information, increasing control over their health care and engaging in health decision-making (Alpay et al., 2009). Especially with regard to chronic conditions, self-management constitutes a crucial asset to improve or at least avoid deterioration of the current state of health. Self-management can include a wide array of actions including reliably taking medication, changing health behaviors and lifestyles as well as performing preventive measures (Fox & Rainie, 2000; Alpay et al., 2009). Moreover, autonomously informing oneself about the disease, potential treatment strategies and prospects of success enable patients to actively participate in patient-doctor interactions (Tonsaker et al., 2014). Increasing the knowledge and control over one's disease also significantly affects how a patient accepts and copes with its consequences in every-day life (Korp, 2006). Secondly, a patient experiences a sense of empowerment due to increased self-help or participating in mutual-help groups leading to perceived personal growth (Staples, 1990; Dickerson, 1998). Three aspects mainly constitute the idea of personal empowerment: the opportunity to rely on oneself and peers instead of solely being dependent on health professionals, the ability to independently make decisions leading to a sense of self-determination and lastly, mutual support, social identification and engagement (Dickerson, 1998).

Online health interventions come in a broad range of offerings all unified in their advantages of being cost-effective and reaching a greater, more diverse population than conventional personal interventions (Barak & Grohol, 2011). Ybarra and Eaton (2005) classify online health programs into passive and active interventions: the former containing educational websites and screening tools, the latter including support groups and online therapy amongst others. Barak & Grohol (2011) advance this classification by introducing four different categories, which in many cases are closely intertwined. *Educational websites*, whether static or interactive, intend to provide users with health-related information to trigger learning and behavioral changes or enhance awareness and self-efficacy (Stout et al., 2001). *Online counseling or therapy* depicts a web-based alternative to face-to-face therapeutic individual or group treatments. Communication mediums comprise video or audio conferences, chats, forums or mails (Barak & Grohol, 2011; Dowling & Rickwood, 2013). *Interactive, self-guided interventions* tend to be prescriptive online programs which assist consumers in managing their health care to achieve improved outcomes in specific health-related issues. Programs often include self-help therapies, education interventions and professional support modules (Dowling & Rickwood, 2013). *Online support groups* offer continuous emotional peer support and feedback and can be found on social media and blogs, or in forums and private chat groups (Barak & Grohol, 2011). Holding true for all four categories, personalized information tailored to individual users' needs proves more effective in achieving positive health outcomes than solely delivering generic, universal information (Lustria et al., 2009). Moreover, health interventions should principally be regarded as supportive and complementary - not substitutable - to in-person support and treatment (Barak & Grohol, 2011).

Educational Websites

In their structured review of online health interventions, Barak & Grohol (2011) observe an essentially greater impact of interactive websites compared to those, that only passively display static information to their readers. Nevertheless, educating readers online through informative posts on a Canadian fertility-related website has significantly increased users' knowledge about fertility and assisted reproductive technology options and, thus, changed common misperceptions about ideal timing of parenthood. The authors found the exposure to the online content to create an immediate effect, which, however, did not succeed to persist over time questioning the long-term efficacy of the intervention (Daniluk & Koert, 2015). Oenema et al. (2001) further investigated the impact of educational websites on increasing users' awareness of nutrition and intake behavior. Using a randomized controlled trial in the Netherlands, they found positive changes in nutritional behavior for the treatment group, implying the suitability of online nutrition education to achieve healthier diets. Performing a meta-analysis of the existing literature, Shahab & McEwen (2009) investigated 11 randomized controlled trials. They concluded that interactive websites facilitating smoking cessations were effective, especially when comparing them to passive information distributed through booklets or emails. Next to health awareness, disease prevention also belongs to the objectives targeted through interactive web-tools. At a university campus in the USA, Wall (2007) tested the efficacy of an alcohol abuse prevention program available via an interactive educational website. In a study of 20,150 randomly assigned individuals, the program proved successful in reducing abusive alcohol consumption among students. Next to receiving interactive feedback about own alcohol consumption behaviors, users also learned about responsible alcohol use and the consequences of abuse (Wall, 2007). Nevertheless, there also exist studies which fail to present a significant positive impact of educational websites: Balsa & Gandelman (2010) studied the implications of web-based interventions on type 2 diabetes patients in Uruguay. Although offering an interactive tool on their website and the ability to virtually communicate with other users and specialists, participants' knowledge, behavior and health outcomes did not change significantly. The authors explained their results mainly with the difficulty many users experienced with navigating the website since the majority already regarded the usage of the website as a challenge in itself. Moreover, they found a connection between having previously used the Internet to extract health information and subsequent active participation in the intervention (Balsa & Gandelman, 2010).

Online Counseling and Therapy

Online therapy sessions are primarily designed to respond to the increasing demand for mental health services. Robinson & Serfaty (2008) showed that users valued Internet therapy as a professional treatment method available without any in-person interaction. Moreover, they discovered that reaching out to students at the University of London via mail to offer online therapy was effective in addressing individuals with eating disorders that would otherwise have not sought out help. In Sweden, online cognitive behavior therapy was found to significantly reduce depression and distress experienced by patients suffering from tinnitus. Nevertheless, the authors found consistency in therapy sessions and continuation of the program to be key conditions for successful interventions (Andersson & Kaldo, 2004). An Israeli field project specifically designed for people in mental crises was successful in providing mental support anonymously through skilled therapists. The project was considered a success because users considering suicide or other harmful acts could seek psychological support, a clinical referral, or just an open ear from this freely accessible online source (Barak, 2007). Notably, users oftentimes find online counseling or therapy offers embedded in a broader interactive online program.

Interactive, Self-Guided Interventions

Interactive, self-guided interventions are often guided by the incentive to support patients in their health-related self-management. Tate et al. (2001) compared a web-based behavioral weight-loss program to a simple weight-loss education website and found the interactive intervention to reach higher results in terms of weight loss among female study participants in the US. Next to general information, participants in the program additionally received a structured schedule, customized therapist feedback and weekly personal touchpoints all leading to the comparably better outcomes. To enhance physical activity among diabetes patients, McKay et al. (2001) tested the effectiveness of an in Internet-based interactive intervention aiming at increasing physical activity self-management. Results indicate that only those participants who used the program regularly experienced significant better outcomes compared to those study participants not using the program. McKay et al. (2001) thus proposes that attention should be shifted to retaining high levels of involvement with interactive online interventions, given that their great potential in improving health outcomes has been proven significant in many studies before. In his study, constant involvement in the program can lower risk of suffering from coronary heart diseases for diabetes patients by 35-55%. To maintain involvement, he suggests testing the efficiency of introducing incentives or committing patients to a minimum number of program visits per week (McKay et al., 2001).

Online Support Groups and Blogs

Online support groups as a form of health intervention predominantly reach positive outcomes. In a study with Australian prostate cancer patients, Broom (2005) investigated the experience the participants (largely already proficient Internet users) had made with online communities and support groups. He concluded that, by providing mental and informative support, the Internet can empower users in taking more control over their state of health and breaking free from their perceived reliability on a specialist. Conducting a survey with 254 chronically ill patients, Kim & Lee (2014) have shown that communicating with other patients online has helped participants to cope with their disease as well as related negative feelings. Moreover, searching for other users' advise as well as sharing own experiences and opinions had a positive impact on medical outcomes over time. Following the results of their content analysis, Willis & Royne (2017; p. 269) state that "online health communities act as informal self-management programs led by peers with the same chronic disease through the exchange of health information". They define online support groups as a cost-efficient, accessible method for users to enhance their self-management of the disease, commit to problem-solving health issues together and improve their health literacy through peer-to-peer support. Smith (2013) finds that especially socially marginalized groups, such as African Americans in the US, become active in social networks to overcome social or health-related disadvantages experienced in their offline every-day life.

In how far these findings which were primarily obtained from studies conducted in the developed countries can be replicated in the African context will be the focus of the remainder of this paper. Whereas all types of online interventions and programs have proven successful in enhancing health outcomes in the Western world, social, cultural and economic circumstances tend to substantially differ and therefore provide a very distinct setting on the African continent. Moreover, HIV/AIDS depict a socially very delicate topic, subject to shame, guilt and fear. Oftentimes persons affected prefer to remain ignorant and avoid the topic instead of actively dealing with the subject or seeking an open dialogue. Nonetheless, some previous literature has found notable results concerning the relationship between the Internet and HIV awareness and prevention across the world.

2.3 The Internet as a Promoter of HIV/AIDS Awareness and Prevention

Through the Internet, people with HIV/AIDS or those at risk of contracting it can access information and find answers to those questions they feel ashamed or hesitant to ask health professionals or educators in person. Moreover, the Internet offers an anonymous source for peer-to-peer support and exchange of experiences and concerns (Gustafson et al., 1999). Many health providers have extended their offers to the Internet, providing access to vast amounts of information concerning prevention strategies, but also covering all facets of coping with HIV/AIDS including medical treatment options, therapies or social support. Moreover, different websites tend to be specially addressed to certain population groups such as adolescents, homosexuals, or ethnic minorities (Kalichman et al., 2002). Today, global websites providing information in countless languages (e.g. that of the Joint United Nations Programme on HIV/AIDS – UNAIDS) as well as national public and private information providers ensure accessibility to free HIV-related information wherever Internet access is available. In testing the efficiency of these web-based interventions on HIV awareness and prevention, two different strands of studies stand out: first, one focusing on sex education and HIV prevention among young adolescents and second, that targeting men engaging in same-sex relations.

In discussions with high school student focus groups in the UK, Goold et al. (2003) already identified the Internet as an accessible way to disseminate sex education to adolescents in 2003. However, he concluded that information should be displayed in an appealing, entertaining manner to increase access and ensure attention of young people. Moreover, interventions should be incorporated into websites that adolescents regularly visit anyways and created in an interactive, imaginative design. Only by ensuring an attractive approach to sexual health education will it be effective to its target group (Goold et al., 2003). In countries where sexual activity still depicts a social taboo subject, sex education conducted through the Internet appears to be especially successful: Lou et al. (2006) undertook a study at high schools and colleges in China, testing the feasibility and effectiveness of online sex interventions including health and service information, therapy and supportive exchange. Knowledge and awareness related to sexually related topics including HIV/AIDS increased significantly within the treatment group. Moreover, students became less liberal towards sex and positively changed their attitude towards contraceptive use (Lou et al., 2006). Nevertheless, in what sense the discussed studies have contributed to the actual prevention of HIV/AIDS or other sexually transmitted infections (STI) is difficult to estimate and was not within the scope of the studies at hand. Opposingly, Hong et al. (2006) discuss the risks originating from uncontrolled Internet use among young college students in China. They did not only not find the frequent Internet users to have acquired better knowledge about HIV/AIDS, but they also identified a positive association between active sexual status and engaging in online risk behaviors such as visiting pornographic websites or distributing sexually related materials themselves. Contrary to previous findings, Hong et al. (2006) indicate that access to Internet alone can result in a "double-edged sword" as it can be either used to autonomously learn about a more responsible sexual behavior or misused to even encourage thoughtless sexual actions among young people.

The Internet has played a paramount importance in combatting the spread of HIV/AIDS among men engaging in same-sex activity in the US. Rhodes (2004) analyzed user profiles and chat room content of a geographically oriented online HIV/AIDS service for men who have sex with men (MSM). He concluded that it was especially the anonymity as well as the extent to which health educators tailored messages to users' needs that helped them to confide their concerns and receive the same benefits as experienced in conventional face-to-face interventions. Moreover, organizations exist that offer HIV/AIDS information to MSM online in real time using instant messaging. Collecting and analyzing these instant messages, Moskowitz et al. (2009) found online counseling through instant messages to be a feasible method to disseminate sexual health information. In a randomized controlled trial, Bowen et al. (2007) further proved online HIV information to play an efficient role in reducing HIV/AIDS risk across rural areas in the US. Specifically, the Internet can become crucial in informing about HIV/AIDS in areas were centers for gay culture are rare and prejudice and discrimination towards MSM still high, both limiting the ability to discuss HIV-related topics with professionals in person (Bowen et al., 2007). In an experimental study in California, Jaganath et al. (2012) evaluated the effectiveness and acceptability of providing peer-to-peer-led HIV prevention interventions on social media. Targeted especially at high-risk groups, including MSM from ethnic minorities, namely Latino and African American men, the authors trained peer leaders in providing culturally sensitive HIV prevention support on social networking sites. They concluded that by creating an environment where MSM can openly express their sexual concerns without experiencing a social stigma, the Internet provides an opportunity for achieving HIV prevention (Jaganath et al., 2012).

Overall, most research focuses on the impact one specific intervention has on increasing health awareness and prevention for a specific patient group. Little has been done to shed light on the effect of the Internet as a whole. In my empirical analysis, I will therefore investigate how the sudden access to faster and more reliable Internet can affect HIV Awareness and Prevention across an entire country. Since, today, there exist internationally accessible websites to all kinds of health-related topics that are translated into a multitude of languages, I expect the arrival of submarine cables at the African coast to act like the previously missing key to this mass of information which has so far been locked for the African populations. Therefore, I seek to investigate whether the sudden arrival of fast Internet access is a sufficient driver of HIV Awareness and Prevention or whether other barriers additionally hinder HIV-related self-management and empowerment in developing countries.

3. Background

3.1 The Internet Infrastructure in Africa

Internet first arrived in Sub-Saharan Africa via analogue telephone lines in the early 1990s. On the national level, a country-wide backbone network transmitted the local Internet traffic through fiber or copper cables, satellite or wireless transmission to the end user (ITU, 2013; Hjort & Poulsen, 2019). Internet access, however, was slow, expensive and limited by national boundaries. While global connectivity substantially increased in most parts of the world from 1990 onwards[2], Sub-Saharan Africa was excluded from this development, persisting in its digital isolation until 2009 (Cariolle et al., 2018). Only then was the submarine cable infrastructure – so far mainly connecting Asia, America and Europe – extended to the African coast. Collectively funded by private investors, multilateral organizations and local governments, the construction of submarine cables and their connection to the already existing backbone networks was a key step towards global Internet access in Sub-Saharan Africa (Schumann & Kende, 2013; OECD, 2014). By 2015, 15 submarine cables had been connected to landing points at all - but two - countries on the African coastline. Located in close proximity to big cities within each country, these landing points established the connection from the international to the national Internet traffic by connecting the submarine cables to the national backbone (Hjort & Poulsen, 2019).

Suddenly providing high-speed and global Internet access, the deployment of submarine cables has had manifold implications for the African continent. First of all, submarine cable arrival has attracted further investments in backbone networks and last-mile infrastructures (The Economist, 2017). Whereas Internet provision has lacked profitability for international service providers for many years, the arrival of submarine cables has reduced prices of international connectivity and thus, made the maintenance of terrestrial networks more profitable for providers (OECD, 2014; Hjort & Poulsen, 2019). Secondly, this significant extension of the digital infrastructure has led to an increase in Internet access and a reduction of transmission costs in many African countries, thus, making the global Internet more accessible to the often financially restricted end users. This is of paramount importance given that in Africa the majority of visited websites – even those containing local content – are hosted overseas (Kende & Rose, 2015). Thirdly, the submarine cable arrival has enhanced the quality of Internet use by increasing connection reliability, speed of transmission and overall capacity, all driving Internet use up (D'Andrea & Limodio, 2019). Fourthly, submarine cable penetration can have substantial impacts on the countries' economies: The increased access to affordable, well-functioning Internet has the potential to increase international trade and foreign direct investments and offers opportunities for the evolvements of new, innovative business sectors and entrepreneurial activity (Hjort & Poulsen, 2019).

Nevertheless, several limitations remain with special regard to the regional barriers to Internet access. The submarine cable-related enhancements of speed and traffic in the terrestrial fiber cable network decrease with distance to the landing point (Hjort & Poulsen, 2019). Moreover,

[2] In the period between 1990 and 2018, 400 fibre submarine cables have been implemented, responsible for more than 99% of international Internet traffic (Cariolle et al., 2018).

the cost of Internet use is determined by the speed and the distance to the backbone infrastructure (Weller & Woodcock, 2013). Consequently, the further away from the terrestrial network, the higher the cost of accessing the Internet and the greater the digital isolation (D'Andrea & Limodio, 2019). Within countries, restrictive regulations and competition among Internet backbone network providers differ according to the respective areas, leading to geographic variations in Internet prices and accessibility (OECD, 2014).

Since each African country runs one national backbone network which is rarely connected to the networks of neighboring countries, spillovers to adjacent states following one country's submarine cable connection are of no concern (Hjort & Poulsen, 2019). This warrants the existence of a specific treatment date for the analysis: the connection of the first SMC at a landing point at Nigeria's coastline in 2010.

3.2 HIV/AIDS Prevalence and Health Care Provision in Nigeria

Counting 202 million people, Nigeria constitutes the most populous country and the highest youth rate on the African continent (World Bank, 2019). Thus, even with a comparably low HIV prevalence rate of 1.4%, the high total population size nevertheless implies a substantial HIV infection number of 1.9 million people (UNAIDS, 2020). In 2018, Nigeria followed South Africa with the second largest HIV epidemic on a global scale (Awofala & Ogundele, 2018). Even though the rate of new infections has declined by 5% between 2010 and 2017, approximately 2/3 of new infections in West and Central Africa were still reported in Nigeria in 2017 (Avert, 2019). Notably, it is women that are disproportionally infected with HIV in Nigeria, depicting 55.56% of those living with HIV in the country. What is more, women were almost twice as likely to newly contract HIV compared to their male counterparts, figures being 26,000 versus 15,000 of new infections in 2018 (UNAIDS, 2020). HIV/AIDS rates tend to vary between states as well as between urban and rural areas, which is likely caused by a range of factors including different levels of education, economic performance and religion. Moreover, cultural differences reinforced by 400 distinct ethnic groups with varying traditions and customs (e.g. female genital mutilation, polygamy, home births) raise the risk of HIV/AIDS transmissions (Awofala & Ogundele, 2018; Avert, 2019).

The Nigerian health care system is considered underdeveloped and fails to provide a sufficient number of HIV tests and counseling services which further reinforces the country's HIV/AIDS crisis and leaves many people in ignorance of their status (Avert, 2019). Despite relatively high spending on health as a share of GDP at 4.6% and overall improvements in health care institutions in recent years, significant shortages remain with focus to inadequate funding, lack of financial or managerial skills and low health access (43.3%) to the Nigerian people (Welcome, 2011). Aspects affecting the health care provision include the striking demographics (55% of population spread across vast rural areas), the below-average quality of drugs available and the high share of private health care providers (70% as opposed to 30% by public institutions) (Welcome, 2011). Although possessing the largest amount of human resources in health (HRH) when compared with other African countries, Nigeria still reports a lack of doctors and other medical staff paired with insufficient educational offers and inadequate distribution and density of professionals across the country (WHO, 2015). Low productivity, ill-equipped work environments, and attrition of medical staff to the big cities or abroad further amplify the modest health care situation. Launched in 2005, the Nigerian Health Insurance Scheme (NHIS) still today only covers 10% of the population. With over half of the population below the poverty line ($1.9 a day), the Nigerian health care delivery system thus denies the most vulnerable people access to affordable basic health services and social risk protection (Aregbeshola & Khan, 2018; World Bank 2019). Political and economic instability, as well as corruption, and lack of strong institutions further limit the quality of accessible Nigerian health care leaving the responsibility of health management and financing to the individuals (Aregbeshola & Khan, 2018).

Both the prevalence of HIV/AIDS within the country and the numerous challenges faced by the Nigerian health care system, give validation to the need of additional health care and information services which are easily accessible and free of charge to the Nigerian population. In 2001 – 2004, a mass media campaign on reproductive health and HIV initiated by the US development agency USAID was greatly received and accepted in Nigeria. Across the country it raised awareness on HIV and educated the population on family planning, contraceptives and HIV prevention methods through television, radio and printed advertisements (Keating et al., 2006). Health information and education provided by private and public, national and international health providers through the Internet could, thus, play a vital part in enhancing Nigerians' health-related preventive self-management and early awareness towards diseases and symptoms. Given that Nigeria's official language is English which is widely used in education, professional interactions and official purposes, the Internet as a universal source of health information gains even more potential.

4. Data

To estimate the impact of fast Internet access on HIV Awareness and Prevention, I exploit the possibility to match the geo-coded household-level data of survey respondents with a terrestrial network infrastructure map. Measuring the distance between households' clusters and the closest points on the network, I can then allocate the individual survey data to this geo-coded information for further analysis. While aiming to investigate the relationship between fast Internet and HIV Awareness and Prevention in Africa in general, I have decided to focus my study on one country solely, Nigeria. For this country, sufficient survey rounds pre and post treatment are available, and all variables of interest are contained in all survey rounds. For my study I have compiled data from two different sources: the individual survey data including the HIV Awareness and Prevention indicators was derived from the Demographic and Health Survey (DHS) while the terrestrial network map originates from the Network Startup Resource Center (African Undersea and Terrestrial Fibre Optic Cables Map – "AfTerFibre").

4.1 Outcome Variables: HIV Awareness and Prevention

The outcome variables are compiled from the Demographic and Health Survey (DHS) Program, which is a comprehensive dataset of nationally representative standardized household and individual surveys conducted in many African countries. DHS survey data has been used as a common data source for numerous research papers in the fields of development economics. The survey asks questions mainly concerning health, labour market participation and nutrition to men and women aged 15-49 and records GPS coordinates for sampling clusters of individual households. From the standard DHS surveys rounds for Nigeria, I use cross-sectional data from the years 2003, 2008 and 2013. Although an additional survey round for 2018 exists, it fails to include several of the key indicator questions linked to HIV leading to my decision to exclude that round. Using Nigeria as the country under investigation has two advantages: first, it contains the most coherent survey rounds including all required HIV-related information in three consecutive turns, and second, geo-coded information of the sampling clusters is available for all three rounds. Next to all required indicators of HIV Awareness and Prevention, I also retrieve data on respondents' characteristics such as age, wealth scores, education levels and literacy rates from the DHS surveys in order to incorporate them as controls. While DHS surveys target both men and women, they set a particular focus on surveying women. Thus, my overall sample size of 115,144 respondents includes 79,953 women and 35,191 men from 904 clusters across the country. All outcome variables are recoded into binary variables taking the value 1 if the associated questions can be answered with "yes" and 0 otherwise.

To examine a broad variety of HIV-related variables from the dataset while reducing the total number of hypotheses tested, I construct two summary indices that measure **HIV Awareness** and **HIV Prevention** and, thus, pool multiple outcomes into two single tests. I build the HIV Awareness Index from three sub-indices that commonly define awareness: **Attitude towards people with HIV, Protection Awareness** and **Transmission Awareness** (Figure 1). In a next step, I then create a **Final Index** combining both, HIV Awareness and Prevention. Given the wide variety of primary outcomes available to predict HIV Awareness and Prevention, this approach aims to define one specific method to weigh the indicators against each other, enabling a more coherent interpretation of results. When testing the primary outcomes separately instead, the risk of obtaining a variety of different results emerges making an unequivocal analysis of results impossible (Berlin et al., 2017).

To construct the different indices, I selected suitable variables of interests from the surveys. For the first sub-index **Attitude towards people with HIV**, I included four indicator variables. The first variable *aidsvendor* takes the value 1 if the respondent would buy vegetables from a vendor who has HIV/AIDS. *Aidsfood* takes the value 1 if the respondent thinks one can contract the disease from sharing food with an infected person. Moreover, *aidsteach* is included, which takes the value 1 if the respondent believes a person with HIV/AIDS should be allowed to continue teaching in front of a school class. These first three variables seek to evaluate the stigma attributed to HIV-infected people as well as the HIV-related discrimination and prejudice existent within a society. Finally, *aidshealthy* takes the value 1 if the person thinks someone may have contracted AIDS although looking healthy. Crucially, lacking the knowledge that a healthy-looking person can have contracted the disease but remains asymptotic implies individuals to expect an externally visible progression of the disease (DHS, 2003). This flawed perception might potentially lead to discrimination towards a more visibly ill patient group and, meanwhile, to an incautious social handling of asymptoticly infected persons. Thus, by including these four variables into the Attitude Index I expect to measure how people think or feel about the disease and whether they express any sense of prejudice or discrimination against people with HIV/AIDS. Crucially, a relatively negative attitude towards HIV-infected people fueled by anxiety and ignorance might hinder the combat against the disease as well as the social acceptance of a public awareness campaign.

The second sub-index, the **Transmission Awareness Index**, seeks to measure in how far respondents have gained an understanding of how the disease is spread and where exactly the risks of transmission lie. Thus, the first three variables *aidspregnant*, *aidsdelivery* and *aidsbreast* take the value 1 if the individual knows that HIV/AIDS can be transmitted from a mother to its child during pregnancy, during delivery or during breast feeding, respectively. Notably, 90% of HIV infections among children below 15 years in Nigeria are caused by mother-to-child transmissions (MTCT) (Chukwuemeka, 2014). In their study on HIV awareness and knowledge among pregnant women in Nigeria, Abiodun et al. (2007) classified knowledge on MTCT routes as inadequate with great room for improvement through large-scale media campaigns. Furthermore, the variable *aidswitch* is included, which takes the value 1 if the respondent thinks one can get HIV/AIDS due to witchcraft or supernatural means. This variable plays an important role in the context of Nigeria, as superstition and the belief in witchcraft both continue to impact logical reasoning of disease causation and consequential health seeking attitudes to date (Archibong et al., 2017). Moreover, indigenous healing ceremonies and traditions oftentimes compensate the lack of accessible conventional medicine, especially in rural areas. Lastly, the variable *aidsbite* takes the value 1 if the respondent thinks you can contract HIV from mosquito bites. The transmission of HIV through mosquito bites is one of the greatest HIV-related misconceptions still persistent, underlining the need for knowledge-enhancing public awareness measures (Bamise et al., 2011; van der Maas & Otte, 2009). Unfortunately, data on knowledge of transmission through drug injection or homosexual intercourse were not contained in the surveys, creating a partly incomprehensive index on transmission awareness.

The third sub-index to constitute the HIV Awareness Index, the **Protection Awareness Index**, aims at investigating the extent to which respondents know the protection measures and understand their significance. The most promising steps to protect oneself and others from an HIV infection include the correct use of a condom, the reduction of the number of sexual partners and regular STI testing. Awareness of these measures is captured by a variety of questions within the Nigerian standard DHS surveys. *Aidspartner* takes the value 1 if the person thinks having only 1 sex partner reduces the risk of contracting HIV. Furthermore, the variable *aidscondom* takes the value 1 if the person thinks always using a condom reduces the risk of contracting HIV. *Justifiedcondom* takes the value 1 if the wife is justified to ask her husband to use a condom if he has an STI. This variable introduces into the index the fact that to promote safer sex, both sexual partners must have a certain degree of awareness of the risks coming along with an STI contracted by one of the parties. Additionally, *aidstestlocation* takes the value 1 if the respondent knows a place where to get an HIV test. Finally, the variable *condomtaught* takes the value 1 if the person believes that students should be taught about condom use in school. Specifically, this variable measures the extent to which the respondent considers a more open approach towards discussing HIV protection methods from an earlier age on to be appropriate. Thus, regarding early-age education on sexual protection as adequate implies a high Protection Awareness of the respondent. Given the bandwidth of variables available to measure the awareness of the most important protection measures, I consider the Protection Awareness Index to be thoroughly and robustly constructed.

By pooling these three sub-indices into the **HIV Awareness Index**, I, thus, seek to gain a holistic overview of the respondents' knowledge and perception with regard to the risks and counteractive measures related to HIV/AIDS. Notably, I do not limit the definition of awareness solely to knowledge about the existence of HIV. Instead, I consider the possession of information about transmission causes, symptoms and protection measures associated with the disease to derive a more comprehensive understanding and hence, awareness of the disease. In how far ignorance is the primary trigger of HIV transmissions while awareness, a positive attitude and adequate information are the first steps towards fighting the continuous spread of HIV/AIDS has been discussed by multiple authors (Gabel & Pearsol, 1993; Ramirez et al., 2002; Sudha et al., 2005).

The second index, the **HIV Prevention Index**, rather targets the respondents' actions taken to avoid a contraction or a further self-inflicted transmission of the disease. Due to a lack of matching indicator variables available in the standard DHS surveys, I limit the construction of this index to two primary variables. First, *aidstested* takes the value 1 if the person has ever been tested for AIDS. Testing for HIV/AIDS helps people who are so far unaware of their status to understand the disease they are coping with and gain access to health care and social support. Moreover, it allows them to prevent a further transmission of the disease to prospective sexual partners. Second, *condomused* takes the value 1 if the respondent has used a condom during the last time, he or she had sexual intercourse. Using a condom counts as the most effective HIV prevention method. Questions asking whether the respondents could actually get themselves condoms and are able to insist to use them during sexual intercourse, were not included in the Nigerian survey rounds or contained too many missing values. Nevertheless, the information contained in the *condomused* variable allows an insight into the availability of condoms and their actual usage to some degree. Despite including the two most effective prevention methods, HIV testing and condom use, the HIV Prevention Index fails to provide a holistic picture of prevention methods available to respondents.

Overall, the **Final Index** incorporates the two outcome measures HIV Prevention and HIV Awareness. It can, thus, be described as the general HIV/AIDS outcome with regard to all the primary outcomes taken into consideration. Despite already including a variety of different variables, the Final Index, nevertheless, falls short in depicting an all-encompassing understanding of HIV-related outcomes. Further aspects affecting Awareness or Prevention, such as additional ways of transmission or questions specifically targeted at the actual actions taken, are missing in the dataset at hand and, thus, cannot be an integral part of this analysis. Nonetheless, especially with regard to measuring Awareness levels, the variables presented appear to create a comprehensive index.

One can interpret the summary indices as weighted means of the different standardized primary outcomes. A paper by Anderson (2008) defines the detailed steps required for the construction of summary indices as follows: First, I switch the signs for the primary outcomes of *aidsfood*, *aidswitch* and *aidsbite*. This way I guarantee that all variables take on the same direction: the value 1 now refers to a positive response in terms of HIV Awareness and Prevention in all cases. Second, for each variable I create standardized replacements by subtracting its control group mean from the primary outcomes and dividing the results by the control group standard deviation. Due to this step, variable outcomes can be normalized enabling a comparable scale. Third, to create the index variables, I sum the weighted average of the standardized primary outcomes with regard to the index groups as defined above. Variables are weighted by the sum of their row entries in the covariance matrix inverses which must be conducted of the standardized variables for each index separately. By weighting the standardized variables, the indices contain the maximum amount of information extractable from the individual outcomes. Fourth, the HIV Awareness Index and the Final Index are created similarly, using all primary variables identified for each of the indices above (Anderson, 2008; Berlin et al., 2017).

Notably, I could have also tested the 16 indicator variables individually. However, obtaining significant results by coincidence would have imposed the risks of interpreting the causal relationships incorrectly (Anderson, 2008). Notably, a Bonferroni correction could have been conducted to adjust for this increased risk of a Type I error while considering the variables independently. Nevertheless, none of these indicators alone could constitute a concept as broad as that of HIV Awareness and Prevention. By incorporating several indicators into each index, I gain the opportunity to develop a meaningful definition of HIV Attitude, Protection and Transmission Awareness and Prevention. Eventually, the objective of this paper is to examine the impact of fast Internet on HIV Awareness and Prevention in the aggregate and not to measure the effect of Internet on the individual indicator values separately.

Figure 1: HIV Index overview

Index Groups							
HIV Awareness						**HIV Prevention**	
Attitude towards people with HIV		**Transmission Awareness**		**Protection Awareness**		**Prevention**	
Aidsvendor	Would buy vegetables from vendor with AIDS	**Aidspregnant**	Knows AIDS can be transmitted from a mother to its child during pregnancy	**Aidspartner**	Thinks having only 1 sex partner reduces AIDS risk	**Aidstested**	Has ever been tested for AIDS
Aidsfood	Thinks one can get AIDS from sharing food with infected person	**Aidsdelivery**	Knows AIDS can be transmitted from a mother to its child during delivery	**Aidscondom**	Thinks always using a condom reduces AIDS risk	**Condomused**	Has used a condom the last time he/she had sexual intercourse
Aidsteach	Believes a person wih AIDS should be allowed to continue teaching	**Aidsbreast**	Knows AIDS can be transmitted from a mother to its child during breast feeding	**Justifiedcondom use**	Wife is justified to ask condom use if husband has STI		
Aidshealthy	Thinks a healthy-looking person can have AIDS	**Aidswitch**	Knows you cannot get AIDS from witchcraft or supernatural means	**Aidstestlocation**	Knows a place to get AIDS test		
		Aidsbite	Knows you cannot get AIDS from mosquito bites	**Condomtaught**	Thinks children should be taught about condom use in school		

4.2 Treatment Variable: Fast Internet Access

To provide a specific location in Nigeria with fast access to the Internet, two factors must interact: the arrival of the submarine cable at Nigeria's coastline and the location's proximity to the terrestrial backbone network. Specifically, to enable fast Internet transmission in Nigeria, the terrestrial backbone network must be connected to the submarine cable which arrives at the landing point on the coastline. Moreover, for an individual's location within Nigeria to then be able to use the fast Internet access, it must live within connection to the terrestrial backbone network based on geographic proximity. For their analysis of the impact of fast Internet arrival on employment, Hjort & Poulsen (2019) used the "AfTerFibre" 2013 map for geodata on submarine cables and terrestrial backbone networks to measure individuals' connectivity. Matching it with the Africabandwith map, they identified and removed backbone segments which were built after the end of 2009 in order to only consider the backbone network that already was in place by the time the submarine cable arrived. Intending to measure the impact of fast Internet specifically due to the arrival of the submarine cable at the Nigerian coast, I adopt their map instead of using the most recent version available at the AfTerFibre website. As several adjustments and expansions have broadened the extent of the backbone infrastructure in Nigeria since the submarine cable arrival in 2010, using the updated version would result in flawed outcomes which could not entirely be traced back to the interaction between the arrival of the submarine cable and the proximity to the existing terrestrial backbone network.

To match the freely accessible AfTerFibre map adapted by Hjort & Poulsen (2019) with the geocoded data of the DHS clusters from all three survey rounds, I use the Geographic Information System (GIS) application QGIS. In this software, I calculate the distances between the individual's cluster locations and the nearest point on Nigeria's backbone network to define control and treatment group (see Appendix A, Figures 8 - 12). In my baseline analysis, I classify respondents whose clusters are less than 500m away from the network as the treatment group and respondents whose cluster locations lie between 500m and 7.5km as the control group.[3] Thus, the distance measured on QGIS defines the dummy variable $Connected_i$ for each individual i's location. Locations further away than 7.5km from the backbone network are excluded from the baseline analysis. Consequently, the total sample of respondents contains 17.75% respondents from the control group and 2.29% from the treatment group.

Table 1 comprises summary statistics of both the outcome variables as well as respondent characteristics included as control variables. Treatment and control group appear to resemble one another in terms of age and gender: on average, respondents are female and 29 years old. Notably, in the pre-treatment period, respondents from the treatment group are more literate, have slightly higher education levels[4] and are more likely to live in urban areas. In the period after the treatment, however, urban location as well as education and literacy levels appear to increase for the control group, even catching up with the connected respondents in terms or literacy. Concerning access to other technical devices, respondents from the treatment group more often used the radio and especially the television as a source of information compared to the control group. Moreover, they are more likely to read a newspaper or magazine at least once a week. After the treatment, however, these rates appear to decrease substantially for both groups – an observation which requires further attention. Thus, to control for these unexpected changes in individuals' characteristics in terms of reduced access to tv, radio and news, I will include all controls discussed into the model specifications to closely monitor individual characteristic differences over time. While Hjort & Poulsen (2019) decided to forego individual-specific fixed effects, I consider them to be of value here in light of the unexpected average decreases.

---

[3] Why 7.5km serves as the threshold for the definition of the control group will be established in section 5.1.

[4] The treatment group reports to have – on average – an incomplete secondary education whereas the control group states a complete primary education.

Table 1: Summary statistics

	Mean Value					Min	Max
Sample	Entire Sample	Control Group	Treatment Group	Control Group	Treatment Group		
Time Period	Both	Pre	Pre	Post	Post		
Controls							
Age	29.3434 (9.8661)	29.3978 (10.1267)	29.7504 (10.0355)	29.2597 (9.5687)	29.1175 (9.6214)	15	59
Gender	0.3191 (0.4661)	0.3146 (0.4644)	0.3349 (0.4722)	0.3207 (0.4667)	0.3295 (0.4702)	0	1
Education	2.9663 (1.665)	2.814 (1.6842)	3.0941*** (1.6629)	3.1011 (1.6272)	3.1327 (1.6657)	0	5
Urban	0.7258 (0.4461)	0.6581 (0.4744)	0.8501*** (0.3571)	0.7658 (0.4235)	0.8753 (0.3305)	0	1
Literacy	0.721 (0.4485)	0.6942 (0.4608)	0.7631*** (0.4254)	0.7448 (0.436)	0.7308 (0.4437)	0	1
TV	1.6908 (1.0564)	1.8953 (1.2171)	2.259*** (1.0226)	1.435 (0.7705)	1.3116 (0.8109)	0	3
Radio	1.8179 (1.0039)	2.1689 (1.0537)	2.3008*** (0.9482)	1.4146 (0.7726)	1.3716 (0.7832)	0	3
News	0.7627 (0.948)	0.8413 (1.0372)	1.0016*** (1.0802)	0.6519 (0.8144)	0.6856 (0.8275)	0	3
Outcome Variables							
Attitude towards HIV	0.0169 (0.7149)	0. 0142 (0.7322)	0.0987*** (0.7036)	0.0018 (0.7051)	0.068 (0.653)	-1.8089	0.8327
Protection	0.0029 (0.6864)	0.0083 (0.6766)	-0.0649** (0.7479)	0.0064 (0.6843)	-0.001 (0.7119)	-2.2347	0.5207
Transmission	-0.0069 (0.8379)	-0.0184 (0.8508)	-0.0333 (0.8411)	-0.0002 (0.8288)	0.0467 (0.8066)	-1.4775	0.7127
Awareness	0.0216 (0.543)	0.0288 (0.5954)	0.0839** (0.5643)	0.0023 (0.4932)	0.0523 (0.4608)	-1.85	1.0079
Prevention	0.1586 (0.9967)	0.1143 (1.0075)	0.2272** (1.0551)	0.1945 (0.9765)	0.1998 (0.9783)	-0.7865	1.8842
Final Index	0.0078 (3.1102)	0.0205 (0.8421)	0.1223** (0.8136)	-0.0417 (4.3331)	0.1583 (4.2221)	-12.8519	9.2292
(%) of Sample	100%	47,8%	5,43%	41,08%	6,01%		
Sample Size	23,076	10,956	1,254	9,479	1,387		

For the pre-treatment period, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ for a t-test between the control group (column 2) and the treatment group (column 3)

5. Methodology

5.1 Empirical Strategy

This paper seeks to analyse the relationship between HIV Awareness and Prevention and the access to fast Internet in Nigeria. Using the arrival of submarine cables as an exogeneous technological shock is a methodology first coined by Hjort & Poulsen (2019) – first published in 2017 - and adapted by Cariolle et al. (2018), Tian (2018) and D'Andrea & Limodio (2019) to measure the impact of fast Internet on employment, firm performance, firm productivity or financial technology and banking. Usually it is difficult to measure the impact of Internet access, as investments in internet infrastructure tend to be specifically allocated to regions where high returns are expected and incrementally improve over time (Hjort & Poulsen, 2019; D'Andrea & Limodio, 2019). In this case, however, I am able to counter this problem by the cross-regional variation in locations' closest distances to the already existing backbone network. Since only the arrival of the submarine cable changes the connectivity to fast Internet access for those individuals living sufficiently close to the terrestrial backbone, I am able to overcome this issue of endogeneity. Moreover, considering the arrival of the first submarine cables in Africa between 2009 and 2012 as an exogenous shock is valid as these cable connections were primarily established as investment objects for private investors with the aim to further increase the Internet connectivity between the US, Europe and Asia (D'Andrea & Limodio, 2019). Hence, this natural experiment permits the identification and testing of a causal relationship between fast Internet access and the constructed HIV indices. In Nigeria, specifically, the first submarine cables to arrive at the country's coastline in 2010 was the MainOne connecting Ghana and Nigeria – and later also Côte d'Ivoire and Senegal – to Europe.[5]

The application of a difference-in-difference model enables me to measure the impact of fast Internet access by comparing the change in HIV Awareness and Prevention experienced by those households that have been connected to fast Internet through the arrival of the submarine cable in 2010 and the change in outcomes for those who have not. As described above, two factors must be guaranteed to provide a specific location with access to fast Internet: First, a submarine cable from overseas must arrive at a landing point at the Nigerian coastline and be connected to the terrestrial backbone. Second, the specific location within the country must be in geographic proximity to the already existing backbone infrastructure, which then delivers the fast Internet. To construct the difference-in-difference model I use two survey rounds to measure the trends pre-arrival (2003, 2008) and one survey round that was conducted post-arrival (2013). I construct a time-variant indicator variable *Arrival*, which signals whether the submarine cable had already arrived in the year of the specific survey rounds. Moreover, I divide the sample into a treatment and a control group based on the smallest distance between the individuals' clusters and the terrestrial backbone network. Hence, for the second, time-invariant indicator variable, *Connected*, I count an individual as connected to the Internet if its cluster is less than 500m away from the closest point on the terrestrial backbone network. Inversely, an individual whose cluster is more than 500m away from the network is considered not connected. The threshold distance of 500m

[5] For a map of the submarine cables connecting the African continent to the global Internet, see Table 13, Appendix A. When arriving at a country's coastline, the submarine cables are connected to the country's terrestrial backbone through landing stations.

is adapted from Hjort & Poulsen (2019) and based on the fact that broadband signals can be transmitted to locations up to 500m away.[6] Notably, spill-over effects of fast Internet access might impact locations further away than the threshold of 500m, an empirical concern taken into consideration in section 7.1 on the first robustness test.

The key identifying assumption of the difference-in-difference model states that individuals from the treatment group and the control group were on parallel trends in HIV Awareness and Prevention rates before the arrival of fast Internet access in Nigeria. Moreover, both groups are assumed to not have experienced "systematically different idiosyncratic shocks" after the fast Internet arrival (Hjort & Poulsen, 2019, p. 11). Given the vast geographical area, including all remaining Nigerian respondents into the control group would neglect the parallel trends assumption. Moreover, it would raise the risk of experiencing idiosyncratic shocks in certain regions more than others. Whereas Hjort & Poulsen, thus, limit the control group to those survey respondents living between 500m and 10km away from the closest network point, I have geographically concentrated the control group even further: Given some substantial differences to more remote survey clusters with reference to the HIV indices and control characteristics pre-treatment, I defined the control group to those respondents living within 500m and 7.5km distance to the closest point on the terrestrial network.

To estimate the main treatment effect of fast Internet arrival, the two indicator variables *Arrival* and *Connected* are interacted (*Arrival * Connected*). The interaction term, thus, compares the change in HIV indices for the connected individuals to the change in outcomes for those individuals not connected. Specification (1) therefore looks as follows:

$$Y_{i,d(i),t} = \beta_0 + \beta_1 Arrival_t * Connected_i + \delta_{d(i)} * Connected_i + \gamma_t + X_i + \varepsilon_{i,d(i),t}$$

where $Y_{i,d(i),t}$ denotes the outcome of one of the six summary indices (Attitude, Transmission, Protection, Awareness, Prevention or Final Index) for individual i in grid cell d(i) and year t. As discussed above, $Arrival_t$ and $Connected_i$ are both dummy variables equal to 1 if the cable has arrived at the Nigerian landing point ($t > 2010$) and if the individuals' cluster is in sufficient proximity to the backbone network (distance < 500m), respectively. Similar to Hjort & Poulsen (2019), I control for district specific, time-invariant differences in the HIV indices by introducing the interaction term $\delta_{d(i)} * Connected_i$. Including this interaction allows me to eliminate the impact of time-invariant factors that might influence the index outcomes and might simultaneously correlate with location, such as factors affecting access to information for more remote areas compared to those areas closer to urban zones. Specifically, $\delta_{d(i)}$ are grid-cell fixed effects accounting for geographical districts of approximately 10x10km (0.1x0.1 degree) square size. Since grid cells resemble clusters in some cases[7] and thus, mostly lack a balanced variation between treated and control group respondents, the interaction of $\delta_{d(i)}$ with $Connected_i$ splits the grid

[6] Hjort & Poulsen (2017) consider 500m to be a conservative proxy of fast Internet connection reach when expecting copper-cables to constitute the last mile technologies. However, knowing that this radius might not depict the actual reach of the Internet access, an extended connection distance will also be considered.
[7] Individual and household geospatial data are collected on the cluster level. Consequently, each cluster is either considered connected to the Internet or not. However, whereas the exact same neighbourhoods (thus, clusters) rarely appear in consecutive survey rounds, the grid cells (separated into treated and untreated cells) remain the same (Hjort & Poulsen, 2019).

cells into treated and untreated districts where applicable. Each grid cell is then either considered connected or not connected – in this regard there exists no variation between the individuals of one grid cell anymore. Thus, separating treatment and control groups for the district specific fixed effects allows to account for the sample not being sufficiently balanced at the grid cell level in terms of treatment status. Additionally, γ_t includes time-period fixed effects to control for differences in HIV index outcomes that are location-invariant. Contrary to the original difference-in-difference model, I therefore do not include the dummy variables $Arrival_t$ and $Connected_i$ into the specification separately. As $Arrival_t$ is already captured by the time fixed effects and $Connected_i$ by the location fixed effects, adding them once more would lead to collinearity. X_i depicts the individual level control variables and includes age, gender, educational levels, residual area (urban or rural), literacy as well as access to television, radio and news.[8] As it is the individuals' cluster locations that are considered in terms of assignment of treatment, I cluster the standard errors at the grid-cell (location) level d(i).

Although Internet is theoretically accessible to anyone within the treatment group, the younger generation is considered to constitute the group of most active online users. Chair & de Lannoy (2018) found 64% of the Nigerian younger age group to possess their own mobile phone (48% a feature or smart phone) – the primary device to access the Internet in Sub-Saharan Africa. Given that HIV awareness campaigns are oftentimes targeted at the younger generation in order to reduce the rate of new transmissions, I also expect the younger part of the treatment group to be more impacted by access to fast Internet in terms of HIV outcomes. I therefore consider the group of 15 to 34-year-olds, specifically, by including an interaction term between $YoungAge_i$ and the arrival of fast Internet for the treatment group in specification (2):

$$Y_{i,d(i),t} = \beta_0 + \beta_1 YoungAge_i + \beta_2 YoungAge_i * Arrival_t + \beta_3 YoungAge_i * Connected_i + \beta_4 Arrival_t * Connected_i$$
$$+ \beta_5 YoungAge_i * Arrival_t * Connected_i + \delta_{d(i)} * Connected_i + \gamma_t + X_i + \varepsilon_{i,d(i),t}$$

Considering the fact that women in Nigeria are disproportionally affected by HIV infections and report almost double the amount of new infections, which have been counted among men in recent years, the question arises, whether men benefit more from access to fast Internet in terms of HIV education and information reception. I therefore extend specification (1) through a gender component to identify whether access to fast Internet has a stronger impact on HIV index outcomes for the male respondents. Specification (3), thus, reads as follows:

$$Y_{i,d(i),t} = \beta_0 + \beta_1 Male_i + \beta_2 Male_i * Arrival_t + \beta_3 Male_i * Connected_i + \beta_4 Arrival_t * Connected_i$$
$$+ \beta_5 Male_i * Arrival_t * Connected_i + \delta_{d(i)} * Connected_i + \gamma_t + X_i + \varepsilon_{i,d(i),t}$$

---

8 Despite the fact that I am using repeated cross-sections data instead of a true panel, I decide to include individual fixed effects. I discussed the reason in section 5.2: Since the individual characteristics, with special attention to access to tv, radio and the news, did not change over time as I would have expected and might therefore affect my outcome variables accordingly, I decide to control for these effects instead. However, keeping in mind that in my model I am not actually observing the same respondents throughout the different survey rounds, respective results with regard to the individual characteristics must be considered with caution. By simultaneously controlling for the grid-cell fixed effects I am nevertheless also accounting for differences in individuals across regions that may bias the index outcomes.

In specification (4) I combine the reasonings of specifications (2) and (3) in order to estimate in how far being male and having access to fast Internet affects HIV index outcomes when also belonging to the young age group. Hence, I divide the sample group and focus on the 15 to 34-year-old respondents only. Given the unequal distribution of Internet-able devices across gender, I also expect a gender-related difference in outcomes even when regarding the youngest, most progressive generation only.

$$Y_{i,d(i),t} = ß_0 + ß_1 Male_i + ß_2 Male_i * Arrival_t + ß_3 Male_i * Connected_i + ß_4 Arrival_t * Connected_i$$

$$+ ß_5 Male_i * Arrival_t * Connected_i + \delta_{d(i)} * Connected_i + \gamma_t + X_i + \varepsilon_{i,d(i),t}$$

Literacy is a key condition to effectively use the Internet as a valuable source of information. It is therefore a crucial next step to focus on the literate respondent subgroup by dividing the sample accordingly. Notably, in Nigeria the literacy rate of men exceeds that of women by far (69.2% vs. 49.7%) on a national level (World Bank, 2019). More than a third of the 15 to 19-year-old girls are assumed illiterate compared to 19.8% of the boys in the same age group. Illiterate rates tend to increase substantially for those girls in the age group that are already married (The Girl Effect, 2016). Consequently, it is considerable, that men – by being more literate - can make more use of fast Internet access and constitute the majority of Internet users in Nigeria. Nevertheless, even when only comparing literate respondents, I suspect a difference in Internet access and thus, HIV outcomes between men and women. Although the female mobile phone ownership rate – and thus mobile Internet rate - has risen in recent years, the one of males still exceeds it by far. Across all low- and -middle income countries worldwide, there continues to persist a 23% gender gap in mobile Internet use (GSMA, 2019). Whether this difference holds in Nigeria and affects HIV outcomes, I seek to find out through specification (5). Therefore, I adapt specification (3) by focusing on the literate respondents within the sample group only:

$$Y_{i,d(i),t} = ß_0 + ß_1 Male_i + ß_2 Male_i * Arrival_t + ß_3 Male_i * Connected_i + ß_4 Arrival_t * Connected_i$$

$$+ ß_5 Male_i * Arrival_t * Connected_i + \delta_{d(i)} * Connected_i + \gamma_t + X_i + \varepsilon_{i,d(i),t}$$

5.2 Internal Validity

The difference-in-difference approach enables the investigation of the effect of a treatment – in this case the arrival of fast Internet – by comparing changes in the outcomes measured between a group that received the treatment (treatment group) and one that did not (control group). In this study, the treatment group consists of those survey respondents within a 500m radius around the terrestrial backbone whereas the control group lives within a 500m - 7.5km radius around the network and is, thus, not expected to experience direct effects from the arrival of fast Internet. The identifying assumption to ensure internal validity of the model and an unbiased difference-in-difference estimator is the **Parallel Trend Assumption**. Precisely, it states that in the absence of the treatment, the difference in outcomes would have remained similar between treatment and control group when comparing the pre- and post-treatment periods. Whereas the assumption does not require actual outcome values to be the same, it expects both groups to follow similar trends had the treatment not occurred. As outcomes in the post-treatment period cannot be observed without a treatment intervention for the treatment group, it is impossible to verify this assumption. Instead, I can test parallel trends by examining the two survey rounds conducted before the treatment (2003 and 2008). If trends were parallel before fast Internet was introduced through the arrival of submarine cables, the DID estimator of the causal effect of fast Internet access on the HIV index outcomes can be assumed unbiased. Only in this case can the deviance between the trend in outcomes of the connected respondents and the trend of the control group be directly associated to the effect of fast Internet access instead of differences in the groups' characteristics (Lechner, 2011).

With regard to the levels of outcomes, differences between treatment and control group in the pre-treatment period can be observed (Table 1). Survey respondents in the treatment group appear to have higher outcomes in the Awareness, Prevention and Final Index than their control group counterparts (compare Table 1, column 2 and 3). Concerning the sub-indices, the respondents from the treatment group reported higher levels for the Attitude Towards HIV Index but lower results for Protection and Transmission Indices compared to the control group in the pre-treatment phase. These outcomes, however, do not risk violating the Parallel Trend Assumption as long as the differences between the treatment and control group outcomes remain equal over the pre-treatment period. Instead, they simply indicate that levels of both awareness and prevention were already higher in the treatment group before the submarine cable arrival. These observations coincide with the comparably better outcomes in terms of control characteristics, i.e. education and literacy, access to tv, radio and news, indicating a more progressive development status of the treatment group respondents a priori.

To test whether the parallel trends assumption holds prior to the treatment in 2010, I graphically verify whether the HIV indices indeed follow parallel paths when comparing the two groups. The pre-Internet arrival trends of the Attitude towards HIV Index are shown in Figure 2. In both survey rounds, the treatment group reports higher index values than the control group. The trends can be considered almost, although not perfectly, parallel. While values are increasing for both groups, the difference between them appears to increase slightly in the second survey round. Notably, the same holds true for the Transmission Index (Figure 3). For both groups, values decrease in the second survey round with trends being almost parallel despite levels slightly approaching each other in 2008.

Figure 2: Parallel trends assumption - HIV Attitude Index

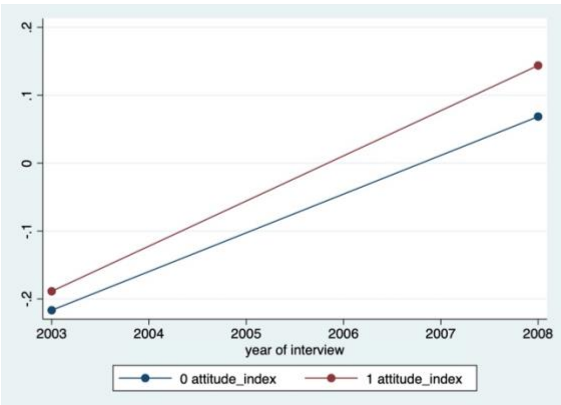
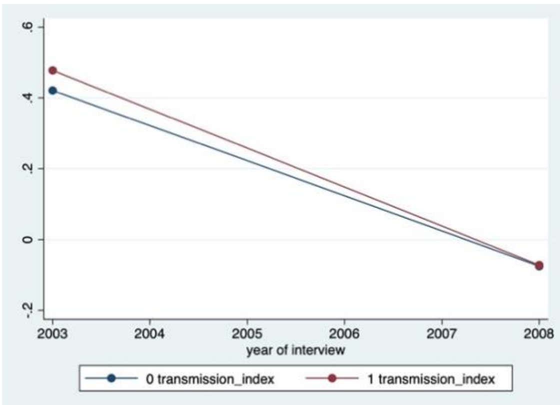


Figure 3: Parallel trends assumption - HIV Transmission Index



For the Protection Index, the Parallel Trend Assumption clearly does not hold (Figure 4). Whereas index values are substantially higher for the treatment group in the year 2003, they drop substantially in 2008 whereas the control group records a slight increase in the index. Consequently, estimator outcomes with regard to the Protection Index must be considered with caution keeping potentially biased results in mind. To what extent this affects the validity of outcomes in this case will be discussed in section 6.1. The Awareness Index does also not demonstrate a clear parallel trend observation (Figure 5). Whereas the levels increase strictly for the control group, the treatment group only registers a slight rise. Again, this index must be considered cautiously when discussing the results in the next section.

Figure 4: Parallel trends assumption - HIV Protection Index

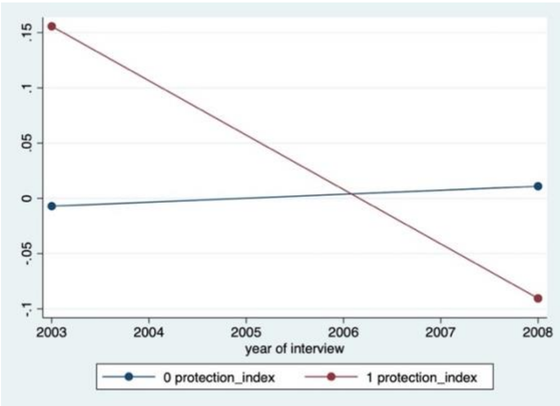
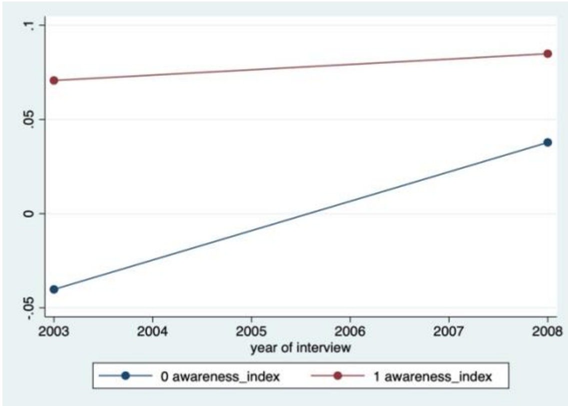


Figure 5: Parallel trends assumption - HIV Awareness



With regard to the Prevention Index, the parallel trend assumption appears to hold to a reasonable extent (Figure 6). Although not perfectly parallel, both groups report increasing index levels with fairly similar slopes. The same can be noted for the Final Index (Figure 7): both groups have almost parallel upward trends when combining all HIV Awareness and Prevention related variables.

Figure 6: Parallel trends assumption - HIV Prevention Index

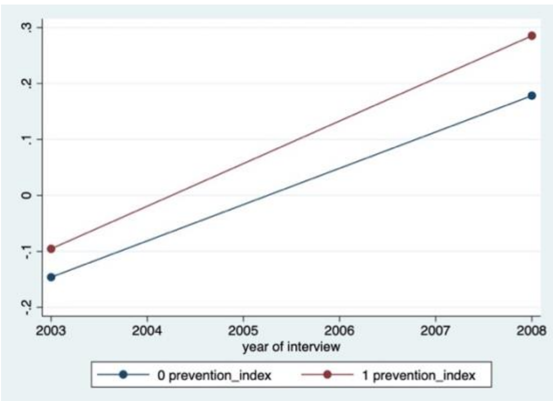
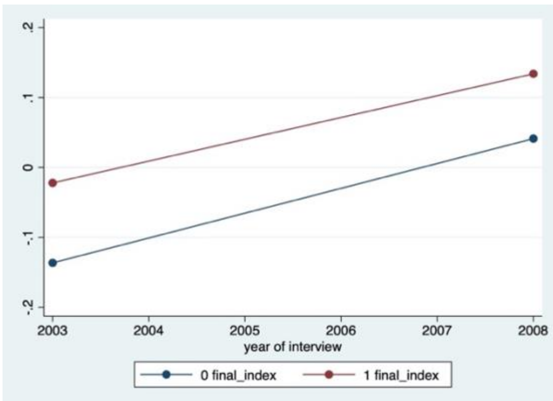


Figure 7: Parallel trends assumption - HIV Final Index



Concluding the parallel trends verification, the assumption does not hold to full extent in all cases: Whereas four out of the six indices outline almost parallel trends, two – Protection Index and Awareness Index – fall short in guaranteeing that the key identifying assumption holds. Thus, in the further analysis of results, outcomes referred to these two indices must be considered with caution and cannot assure to demonstrate an unbiased effect.

In his discussion of the difference-in-difference estimation strategy, Lechner (2011) states three further assumptions required to identify a mean causal effect. The first is the **Stable Unit Treatment Value Assumption (SUTVA)** which implies that assigning the treatment to one respondent should not affect the outcome for other respondents. Thus, there should be no interference in results between control and treatment groups based on the treatment assignment for the latter. Consequently, a spillover effect of the Internet access from the connected area to its surroundings – the respondents of the control group, specifically – would harm this assumption. The probability that the availability of fast Internet access in the treatment group also indirectly impacts HIV Awareness and Prevention in the control group is not entirely unrealistic as the considered distances are relatively small. As respondents from control and treatment group theoretically live fairly close together, those living too far off the network to access fast Internet from home might regularly carry their mobile phones or other Internet-ready devices to areas closer to the network to connect to the Internet. However, given that connection to fast Internet deteriorates strongly within short distance from the backbone network, this might only be applicable for those within the control group that live very close to the threshold radius. In order to further test the steadfastness of this assumption, I include a robustness test in section 7.1 to test the effects of fast Internet access using a different treatment threshold. The corresponding results confirm my expectations indicating that effects are reduced in magnitude and lose statistical significance when increasing the treatment threshold from 500m to 1km. Consequently, SUTVA does not appear to be strongly violated.

The second assumption – the **Exogeneity Assumption** – states that the individual control variables contained in X_i may not be affected by the treatment effect (Lechner, 2011). The individual characteristics included are age, gender, literacy rates, education levels, residual area as well as access to tv, radio and newspapers. The first five variables are individual characteristics that I consider not to be influenced by the arrival of fast Internet. I include access to tv, radio and newspapers to control for alternative channels through which individuals might obtain information on HIV Awareness and Prevention measures. However, I also do not expect access to any of these sources to be directly affected by the arrival of fast Internet. Due to the Exogeneity Assumption I have excluded a control variable measuring the wealth index across the respondents. As the previous papers using the same methodology have shown, fast Internet access can have a substantial impact on firm performance, increased market competition, innovative entrepreneurial spirit and employment, all drivers of increased wealth (Cariolle et al., 2018; Tian, 2018; Hjort & Poulsen, 2019; D'Andrea & Limodio, 2019).

The final assumption states that the treatment cannot already influence the respondents in the pre-treatment period. Thus, prior to the treatment the treated group should not already depict changes in their behavior associated with the treatment anticipation (Lechner, 2011). As the arrival of fast Internet access in 2010 cannot increase or decrease HIV Awareness or Prevention in the years prior, this assumption is most likely not violated. There is no evidence that respondents in the treatment group would consciously await the arrival of fast Internet to only then seek to find out certain HIV-related information. Moreover, it is not obvious that respondents in the treatment-group previously knew they would gain access to fast Internet just given their close proximity to the network.

6. Results

This section seeks to describe the results obtained from the empirical analysis. It first focuses on the Awareness Index including its three sub-indices Attitude towards HIV, Protection and Transmission, and then continues with the results for the Prevention Index. Eventually, the empirical results for the Final Index will be presented. Each table is dedicated to one index including the estimators of interest from the five different specifications discussed above. While specification (1) depicts the original difference-in-difference model, the remaining specifications expand it by including additional interactions and focusing on specific sub-sample groups: Specification (2) focuses on young respondents whereas specification (3) addresses the male subgroup. Specification (4) is then a combination of the two, investigating the impact of being both young and male on the different HIV indices. Next, specification (5) regards the effect of being literate and male when measuring the impact of fast Internet on HIV outcomes.

6.1 HIV Awareness Index

To consider the levels of HIV Awareness among the survey respondents, I will first examine the sub-indices individually before considering the results of the combined Awareness Index.

Attitude towards HIV Index

In the first difference-in-difference model specification, the interaction term between the treatment (*Arrival*) and the treatment group (*Connected*) is considered the estimator of interest. Against my expectations, however, it has a negative and insignificant coefficient when considering the Attitude towards HIV Index as the outcome variable (Table 2). Thus, the result indicates that being connected to fast Internet access after the arrival of submarine cables does not significantly change the attitude respondents have towards people affected with HIV.

When turning to the additional interactions and sample alternations, the results also deviate from the expectations. The young age group, in general, appears to have a significantly higher (more positive) attitude towards people with HIV when compared to older respondents. No significant changes can be found in the outcome variable with regard to the main estimator of interest in specification (2): Attitude towards HIV does not appear to be affected by having access to fast Internet as a young respondent.

A similar picture emerges for specification (3). According to the results, male respondents tend to hold a better attitude towards people with HIV than women a priori. Nevertheless, this difference cannot be explained by access to fast Internet: Male, connected respondents do not report a significant change in the index outcomes after fast Internet arrival. Specifications (4) and

(5) reach the same conclusion. When regarding the young (4) or literate (5) subgroup, the male and connected respondents do not significantly change their attitude towards HIV after the submarine cable arrival. Interestingly, however, literate, male respondents tend to hold a better attitude towards HIV affected people than literate women. Again, this difference seems to have another origin than access to fast Internet.

Consequently, there appears to be no significant effect of connection to fast Internet access on Attitude towards HIV for any of the restricted sample groups. In addition, the overall HIV Attitude trends over time provide an interesting insight (see Appendix B, Figure 14): After parallelly increasing from 2003 to 2008, the Attitude Index appears to decrease for both, control and treatment group, again in a parallel trend between 2008 and 2013. Thus, it appears that HIV Attitude is indeed an outcome that might be affected by external influences more than by fast Internet access and for both groups equally – a consideration further developed in the subsequent discussion section.

Table 2: Attitude towards HIV Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0156 (0.0897)	-0.0392 (0.0803)	-0.0111 (0.1023)	-0.031 (0.1273)	0.0406 (0.1122)
Young		0.099*** (0.0253)			
Young * Arrival		-0.0537** (0.0235)			
Young * Connected		-0.0589 (0.0388)			
Young * Arrival * Connected		0.031 (0.0568)			
Male			0.0428* (0.0252)	0.0469 (0.0288)	0.067*** (0.0251)
Male * Arrival			0.036 (0.0497)	0.0411 (0.0517)	0.0133 (0.0522)
Male * Connected			0.005 (0.0514)	-0.0304 (0.0597)	-0.0318 (0.0458)
Male * Arrival * Connected			-0.0142 (0.1008)	0.0068 (0.1132)	0.0234 (0.1031)
N	21,383	21383	21383	14,727	15,870
R_2	0.1728	0.1737	0.1730	0.1684	0.1512
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

HIV Protection Index

Turning to the Protection Index next, I seek to estimate the effect of fast Internet access on respondents' knowledge of protection measures and their significance (Table 3). Recalling the failure to provide evidence for common trends in the period before the submarine cable arrival (compare section 5.2 on Methodology), the outcomes linked to the Protection Index should be considered with caution. When regarding the overall index trends, however, another picture emerges: Whereas the index of the control group first increases slightly to then decrease mildly after the treatment period, the treated group reports a strong fall in the index measure followed by an equally strong rise in the period after the treatment (compare Appendix B, Figure 15). Consequently, both groups experience a break in their trends. The improvement observed in the outcome measure of the treatment group following the submarine arrival is not a continuation of the previous trend. Instead, a factor between the last two period appears to substantially change the protection awareness of the connected respondents to the positive. Consequently, I can relax the common trend assumption in this case given the discontinuity of outcomes altogether. In the following, I will analyse in how far access to fast Internet constitutes this change in direction for the index outcomes of the treatment group.

Specification (1) yields a positive Protection summary size change for respondents living within the 500m radius around the terrestrial network after the submarine cables are connected. However, as the coefficient is not significant, fast Internet arrival cannot be considered to affect Protection Awareness for all respondents within the treatment group.

Once more, I observe significant effects on the Protection Awareness Index for being young or male in all following specifications. For specification (2), however, I cannot find a significant increase in the index for those respondents who are young and simultaneously connected to fast Internet. The sample group adjustments (3), (4) and (5) provide different results: In specification (3) I find that belonging to the male sub-group and gaining access to fast Internet increases the awareness on HIV protection by a coefficient of 0.217, though, an effect only significant at the 10% level. Multiplying this coefficient by the index' sample standard deviation indicates the size of the impact on the HIV Protection Awareness Index expressed in standard deviation of the respective dependent variable.[9] This impact accounts to 0.149 for those male respondents connected to fast Internet (3). The results are replicated when considering only the young (4) or the literate (5) respondents within the sample: Here, after the arrival of the submarine cables, male, connected respondents experience a 0.213 (0.31) and a 0.132 (0.193) effect change in protection awareness, respectively – the latter again only significant at 10%. Consequently, the young and male sub-group (4) experiences the comparably greatest and also statistically more significant effect from accessing fast Internet. Contrarily, in the same specification the entire young sub-group combined experiences a decrease of -0.09 (-0.1317) in the Protection Index through the arrival of fast Internet access, however, only marginally significant again.

9 Here as well as in the following, I present the actual effect size on the index outcome (i.e. the coefficient multiplied by the index' standard deviation). In parentheses I will state the respective coefficient from the regression table. Standard deviations of the indices are as follows: Attitude towards HIV (0.7149), Protection Index (0.6864), Transmission Index (0.8379), HIV Awareness Index (0.543), HIV Prevention Index (0.9967) and the Final Index (3.1102).

Thus, while specifications (1) and (2) do not deliver on the expectations entirely, the remaining three models support the effect predictions to some extent. Consequently, despite lacking a significant positive change in the Protection Awareness Index due to fast Internet arrival for the exhaustive treatment group, there definitely exists a significant effect of the Internet for the young, literate and male subgroups. Since results for specifications (3) and (5) are only significant at the 10% level, however, a definite conclusion on the impact of Protection Awareness can only be drawn for specification (4).

Table 3: HIV Protection Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0113 (0.0481)	-0.0052 (0.0544)	-0.0614 (0.0668)	-0.1317* (0.0706)	-0.0941 (0.0743)
Young		0.1103*** (0.0213)			
Young * Arrival		0.0222 (0.0204)			
Young * Connected		-0.0284 (0.0383)			
Young * Arrival * Connected		0.019 (0.0596)			
Male			0.1419*** (0.038)	0.1572*** (0.0387)	0.1201*** (0.038)
Male * Arrival			-0.0401 (0.0503)	-0.0412 (0.054)	-0.0228 (0.0532)
Male * Connected			-0.028 (0.0888)	-0.0795 (0.0872)	-0.0009 (0.0727)
Male * Arrival * Connected			0.2168* (0.1247)	0.31** (0.1303)	0.1926* (0.1138)
N	20,783	20,783	20,783	14,317	15,369
R_2	0.1608	0.1627	0.1617	0.1686	0.1175
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

HIV Transmission Index

With the third sub-group index I seek to measure the extent to which the respondents have understood how HIV/AIDS is spread and where exactly the risks of transmission lie. Once more, I fail to estimate a significant impact of fast Internet arrival on the index for the complete treatment group (1). Moreover, in none of the other specifications do the difference-in-difference estimators prove to be significant (Table 4). Regarding specification (2), I find a positive change in the Transmission Index for the young subgroup once more. With an effect size increase of 0.4 (0.048), the impact is significant at the 10% level, implying that young people are more aware of transmission routes. Turning to specifications (3) and (4), the results depict a significant decrease in the Transmission Index for all male (and literate) respondents. Thus, whereas – compared to women - the male subgroup reports higher Protection Awareness, it simultaneously experiences less Transmission Awareness. One possible explanation is that the variables included into the Transmission Index mainly concentrate on MTCT. Given that pregnancy, birth giving, and breastfeeding are still oftentimes considered women's' issues, the male sub-group under investigation in specification (3-5) might keep these subjects at distance in every-day life.

The actual estimators of interest, however, pose no significant changes in HIV Transmission Awareness. Consequently, access to fast Internet does not appear to impose an effect on the Transmission Index, neither for the entire treatment group, nor for any of its restricted subgroups.

Table 4: HIV Transmission Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0324 (0.0624)	0.0269 (0.0937)	-0.0469 (0.0728)	-0.0834 (0.0631)	0.0169 (0.0672)
Young		0.0479* (0.0262)			
Young * Arrival		0.0207 (0.0207)			
Young * Connected		0.0378 (0.0444)			
Young * Arrival *Connected		-0.0886 (0.0684)			
Male			-0.0976** (0.0409)	-0.0962** (0.0447)	-0.0495 (0.0347)
Male * Arrival			0.0333 (0.0456)	0.0333 (0.0477)	0.0182 (0.0396)
Male * Connected			0.049 (0.067)	0.1 (0.0741)	0.0551 (0.0727)
Male * Arrival * Connected			0.0437 (0.0923)	0.0127 (0.0988)	0.0006 (0.0935)
N	20,557	20,557	20,557	14,113	15,293
R_2	0.1163	0.1167	0.1166	0.1155	0.1020
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Combining the variables included in the sub-indices into the overall Awareness Index, I obtain the results presented in Table 5. The baseline difference-in-difference model in specification (1) obtains a positive, insignificant coefficient for the estimator. Consequently, I cannot claim that fast Internet arrival changes HIV Awareness. Once more, only the respective subgroups provide expected outcomes: As I estimated in section 5.1, young as well as male respondents are more aware of HIV and its implications than older or female individuals, other things equal. However, the access to fast Internet after the submarine cable arrival at the Nigerian coast does not result in significant changes for the treatment group in any of the specifications considered. Nevertheless, given the lack of parallel trends between treatment and control groups before the submarine cable arrival, results linked to the HIV Awareness Index must be considered with caution anyhow. Thus, the coefficients presented in Table 5 might be biased and therefore deviate from the actual effect of fast Internet access on HIV Awareness. While keeping this in mind, the conclusion of the results at hand still indicate that access to fast Internet does not impact HIV Awareness. The overall trend in the Awareness Index between survey rounds 2008 and 2013 supports these findings: Regarding the treatment group, the index decreases in size despite the arrival of fast Internet implying that its outcome might be impacted by different factors (see Appendix B, Figure 16).

Table 5: HIV Awareness Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0208 (0.0641)	0.0294 (0.0692)	-0.0043 (0.0697)	-0.0534 (0.0809)	-0.0121 (0.071)
Young		0.1036*** (0.0169)			
Young * Arrival		0.0202 (0.015)			
Young * Connected		0.0113 (0.0242)			
Young * Arrival * Connected		-0.018 (0.0358)			
Male			0.0447** (0.0202)	0.0626*** (0.0231)	0.06*** (0.0182)
Male * Arrival			0.0031 (0.0368)	-0.0034 (0.0386)	-0.0091 (0.0354)
Male * Connected			0.0174 (0.0602)	-0.0004 (0.0603)	0.0302 (0.0486)
Male * Arrival * Connected			0.0778 (0.0968)	0.0926 (0.1003)	0.052 (0.0845)
N	20,154	20,154	20,154	13,864	15,004
R_2	0.2243	0.2280	0.2249	0.2258	0.1526
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

6.2 HIV Prevention Index

The Prevention Index measures the respondents' actions taken to avoid a contraction or self-inflicted transmission of the disease. However, lacking sufficient input variables linked to HIV Prevention, the index fails to provide a holistic picture of effective prevention methods – a barrier, which must be considered when evaluating the result obtained in Table 6.

The difference-in-difference estimator does not yield significant results in specification (1) once again. Thus, fast Internet access does not appear to significantly change the Prevention Index when considering the overall treatment group. Concerning the actions taken to avoid an HIV contraction, young respondents seem to perform worse than older ones, while men report taking more action than women. Additionally, there appears to be an increase of 0.068 (0.069) in the Prevention Index for all young respondents – no matter treatment or control group – after the arrival of fast Internet in 2010 (2). Young individuals who live within the 500m treatment radius further report having taken more prevention measurements than those who live more remotely, no matter the time period. Again, there is no direct change in the Prevention Index affected by the arrival of fast Internet access specifically for those who receive the Internet connection. Instead, it appears that young individuals have simply increased their prevention measurements over time independent of Internet access (2).

Specifications (3), (4) and (5) provide similar observations: Whereas the prevention actions conducted by male respondents appear to decrease after 2010 in all cases, the male respondents within the treatment group report higher index outcomes when considering all years combined. Specifically, being male and connected does increase the Prevention Index by 0.16 (0.161) and 0.307 (0.308) considering the entire sample (3) and the young sub-sample (4), respectively. This change, however, stems only from living within the treatment radius and is not directly linked to the actual arrival of fast Internet in 2010. Given that treatment group members are reportedly slightly more educated, more likely to live in urban areas and have more access to other news sources like tv and radio, these differences observed in Table 6 might have a different origin.

To conclude, there is no substantial change in the Prevention Index found that can be directly traced back to the arrival of fast Internet in Nigeria. Whereas certain respondent characteristics appear to impact the index, these cannot be further explained by the ability to gather information online.

Table 6: HIV Prevention Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0893 (0.083)	0.0125 (0.1041)	-0.0396 (0.0726)	-0.0714 (0.1129)	-0.0709 (0.0817)
Young		-0.1057*** (0.0381)			
Young * Arrival		0.0686** (0.0427)			
Young * Connected		0.133* (0.0796)			
Young * Arrival * Connected		-0.168 (0.1053)			
Male			0.4256** (0.037)	0.5744*** (0.0464)	0.5331*** (0.039)
Male * Arrival			-0.0846** (0.0419)	-0.0695 (0.0612)	-0.1102** (0.047)
Male * Connected			0.1606** (0.0707)	0.3081** (0.1226)	0.1324 (0.0843)
Male * Arrival * Connected			-0.1424 (0.1013)	0.0926 (0.1003)	-0.0739 (0.1219)
N	15,467	15,467	15,467	9,378	10,456
R_2	0.1954	0.1961	0.1962	0.2405	0.1910
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.3 Final Index

By combining all variables into the Final Index, the specifications measure the HIV outcomes altogether – incorporating both Awareness and Prevention of the disease. Table 7 represents the results for the Final Index. Notably, all six specifications report a considerably low R_2. This implies that the models at hand perform poorly in explaining the variations within the Final HIV Index through the included explanatory variables. The reason is most likely the vast variety of variables measured by the Final Index. Combining all 16 primary indicators included in the previous indices, the broadness of the index makes it complicated to pin down explanatory variables that can affect all indicators combined in the index equally.

I do not find evidence that access to fast Internet impacts the index in any of the six specifications considered. The coefficients of interest, the interaction term between the submarine arrival and the treatment group in specification (1) and the triple interactions in the remaining models, yield statistically insignificant results in all cases. Thus, not only the entire treatment group, but also all distinguished sub-groups including the young, the male, and the literate respondents do not experience significant changes in the Final Index based on fast Internet arrival. In summary, combining all variables into one index, its effect size realizes no changes due to access to fast Internet for the connected samples group.

Table 7: Final Index

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.4009 (0.5476)	-0.229 (0.5169)	-0.3975 (0.5883)	-0.3631 (0.6182)	-0.3998 (0.6516)
Young		0.2744*** (0.0964)			
Young * Arrival		-0.375*** (0.1126)			
Young * Connected		-0.0259 (0.0822)			
Young * Arrival * Connected		-0.258 (0.2219)			
Male			0.1124** (0.0495)	0.1629** (0.0625)	0.0253 (0.0459)
Male * Arrival			0.0079 (0.0419)	-0.0571 (0.2357)	0.1091 (0.1945)
Male * Connected			0.0828 (0.0894)	0.0395 (0.1208)	0.1779* (0.1032)
Male * Arrival * Connected			-0.001 (0.3915)	-0.1473 (0.4516)	-0.0648 (0.4153)
N	13,981	13,981	13,981	8,460	9,719
R_2	0.0716	0.0730	0.0716	0.0856	0.0672
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

7. Robustness

In the following, I will adapt the baseline estimation of my empirical analysis in order to test the robustness of the methods applied. As a baseline threshold I have defined the Internet connection to reach a distance of up to 500m from the backbone network. Thus, the treatment group contains those households within a 500m radius around the country's network whereas respondents of the control group live within a 500m to 7.5km distance. Specifically, I will discuss three robustness tests, all based on the analysis of Hjort & Poulsen (2019). First, I will vary the treatment radius to confirm whether the effect of fast Internet is highly localized. Second, I will vary the size of the grid cells used to determine the location-specific fixed effects. Specifically, I seek to examine whether the effect of Internet access is correlated to the grid-cell sizes. Third, I exclude more remote locations from the control group by reducing the control radius from 7.5km to 5km.

7.1 Varying the Treatment Distances

For the first robustness test, I divide the sample into treatment and control groups based on a different Internet connection distance threshold than in the baseline estimation. The objective is to examine whether the effect of fast Internet access is sensitive to the distance from the terrestrial network, which is used as the radius to separate treatment from control group. According to Hjort & Poulsen (2019), one can expect a household to be connected to the Internet if it is located a maximum of 500m away from the backbone network. Nevertheless, it cannot be precluded that the effect of fast Internet access extends this radius to some extent. Therefore, I conduct the empirical analysis once more, this time using an extended distance of 1km as the new radius. Now, survey respondents living up to 1km away from the network constitute the treatment group while those in-between 1km and 7.5km make up the control group. Replicating the specifications for the different groupings, I can then determine whether the effect of fast Internet access exists beyond the 500m distance. If this is not the case, I expect the outcome estimators to diminish in magnitude and statistical significance given that the new treatment group now constitutes former members of the control group.

Due to the level of complexity introduced by the re-construction of the indices and the modelling of different specifications, I retain from replicating the analysis 16 times with Internet connection radiuses ranging from 400m to 2km as has been done by Hjort & Poulsen (2019). Instead I focus on one replication, solely, choosing a radius twice the size from the baseline distance. Notably, I will therefore not be able to observe an incremental change in the indices as the connection radius rises – instead, I will only be able to observe a second static snapshot of results. Nevertheless, given I am choosing a relatively large second radius doubling the previous distance, I expect to observe a clear difference in coefficient outcomes. In addition, I will only present the variables of interest – the effect of fast Internet access and its interactions with the different subgroups.

Table 8 shows the replicated results for the HIV Awareness Index. In all specifications the actual coefficients of interest are consistent with the baseline estimations in that no statistically significant effects can be found. Thus, for the adapted treatment group there is also no effect of fast Internet access found, no matter the respective sub-group interaction. Consequently, all specifications pass the robustness test by retaining the non-existent treatment effect on HIV Awareness even for the extended treatment group.

Table 8: HIV Awareness Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0646 (0.0514)	0.0702 (0.05)	0.0467 (0.059)	0.0154 (0.0654)	0.0108 (0.0588)
Young * Arrival * Connected		-0.0109 (0.0303)			
Male * Arrival * Connected			0.056 (0.0596)	0.0676 (0.0625)	0.064 (0.0572)
N	20,154	20,154	20,154	13,864	15,004
R_2	0.2225	0.2261	0.2227	0.2239	0.1512
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

The Prevention Index yields slightly different results (Table 9). The initial model (1) retains the conclusion that access to fast Internet does not affect HIV Prevention for the extended treatment group. Specifications (3), (4) and (5) show identical results: the arrival of fast Internet does not provide an effect on HIV Prevention for either the male, male and young and male and literate sub-groups. Only specification (2) reports different outcomes when using the extended treatment group. Here, the effect of being young and connected to fast Internet also reduces in magnitude, but the p-value is slightly reduced resulting in the respective coefficient now significant at the 10%. Thus, fast Internet access now results in a – although smaller – decrease in the Prevention Index given affiliation with the young and connected part of the sample group. Thus, only for specifications (1), (3), (4) and (5) does the expectation hold, that coefficients decrease in magnitude and decrease (or maintain) their statistical (in)significance given the expansion of the treatment radius. Concerning specification (2) instead, the effect of fast Internet access on the young respondents does not appear to be highly localized.

Table 10 indicates the adapted results for the Protection Index which obtained the most meaningful results in the baseline estimation. In specification (2) – (5), the coefficients of interested have all decreased in magnitude in all cases and have turned from significant at 10% to statistical insignificance in specifications (3) and (5). Notably, while young men connected to fast Internet (4) experienced a 0.213 (0.31) change in protection awareness, this effect was still significant but reduced to 0.055 (0.0795) for the extended treatment group. Only the difference-in-difference estimator in specification (1) does not change as expected: Through the increase in the treatment radius, the effect of fast Internet access on the entire treatment group gains significance at the 10% level. Therefore, I conclude that with regard to the Protection Index, the effect of fast Internet access proves to be highly localized and diminishes with increasing distance from the backbone network for all but the first specification.

Table 9: HIV Prevention Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0081 (0.0609)	0.1006* (0.054)	0.0299 (0.0596)	-0.0049 (0.0826)	-0.0183 (0.0745)
Young * Arrival * Connected		-0.1549* (0.0822)			
Male * Arrival * Connected			-0.068 (0.0799)	-0.1933 (0.1595)	-0.0073 (0.0921)
N	15,467	15,467	15,467	9,378	10,456
R_2	0.1967	0.1975	0.1978	0.2437	0.1918
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Table 10: HIV Protection Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival x Connected	0.1021* (0.0556)	0.0693 (0.0599)	0.0856 (0.0679)	0.0547* (0.0733)	0.0688 (0.073)
Young * Arrival * Connected		0.0459 (0.0471)			
Male * Arrival * Connected			0.0487 (0.0895)	0.0795** (0.0918)	0.0163 (0.09)
N	20,783	20,783	20,783	14,317	15,369
R_2	0.1600	0.1620	0.1602	0.1664	0.1161
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Results obtained for the remaining three indices, Attitude towards HIV, Transmission and the Final Index, can be found in Table 17, 18 and 19 in Appendix C. Here, results resemble one another when comparing the different specifications between the baseline estimations and the extended treatment group outcomes. For all three indices, I cannot claim a significant effect of fast Internet access once more, implying all specifications behave as expected under the increased treatment group. Thus, to conclude, I consider the models to be robust to changes in the treatment radius for all indices - with two exceptions for specification (2) of the HIV Prevention Index and specification (1) of the HIV Protection Index. Consequently, linking back to the model assumptions discussed in section 5.2, I consider SUTVA to hold since a meaningful violation through potential Internet spillover effects from the connected area to its surroundings cannot be found.

7.2 Varying the Grid Cell Sizes

The grid cells used to determine location fixed effects in the baseline estimation are 0.1x0.1 decimal degrees, which translate into 10x10km squares. To test whether the estimation results are robust to the size of the grid-cells, I vary them to 15x15km, 20x20km, 25x25km, 30x30km, 35x35km, 40x40km, another robustness test inspired by Hjort & Poulsen (2019). The intuition is that effects should not be sensitive to changes in the grid-cells. The grid cells are primarily included to account for time-invariant location fixed effects and to cluster the standard errors at the location level, hence they are responsible for eliminating district-specific differences. Increasing the raster dimensions of the grid cells, by which I allocate survey respondents to a specific connected or unconnected location, should therefore not affect the HIV index outcomes. Especially when considering smaller changes in grid cell sizes, I expect regression results to not modify substantially.

Given the scope of this paper, I limit the number of models investigated to one specification per index using the range of new grid cell sizes.[10] Moreover, the regression tables are again limited to presenting only those coefficients considered to be most informative for the purpose of this study – the interaction terms estimating the effect of fast Internet access either on the entire treatment group or defined sub-groups. Specifically, I observe changes in HIV Awareness results for young respondents (specification (2), Table 11), changes in the HIV Prevention Index for the male subgroup (specification (3), Table 12) and changes in the Protection Index for the young and male respondents (specification (4), Table 13). Moreover, in Appendix D, I show the altered outcomes for Attitude towards HIV, HIV Transmission and the Final Index considering the initial specification (1) including the full sample group, the male sub-group (3), and the literate, male subgroup (5), respectively (Table 20, 21 and 22).

When investigating the 2nd specification to measure HIV Awareness by using the different grid cell sizes (Table 11), the expectations are met: the coefficient measuring the effect of being young and connected to fast Internet maintains its magnitude and its statistical insignificance. The

[10] As results for the coefficients of interest are insignificant for all indices but the HIV Protection Index, I have allocated the indices to the specifications randomly. Only the Protection Index I purposefully matched with specification (4) to test whether the changes in the grid cell size affect the most meaningful outcome of the results section – the positive impact of fast Internet access on the male and young respondents.

same holds true for the coefficient measuring the impact of fast Internet access for the entire treatment group. Turning to the Prevention Index (Table 12), the effect of being male (3) and connected to fast Internet access alters from being statistically insignificant to being significant at the 10% for two specific grid cell sizes. Similarly, the effect of fast Internet access for the entire treatment group changes to being statistically significant at the 1% level, implying that when considering the 15x15km grid cells instead, the access of fast Internet has a negative effect on HIV Prevention of the magnitude 0.128 (0.128). Although the remaining coefficients maintain their statistical insignificance, they change signs and turn positive when increasing the grid. Consequently, while the outcomes for the HIV Awareness Index appear to not be sensitive to changes in grid cell sizes (when considering specification (2)), HIV Prevention results (for specification (3)) change slightly in significance with changes in grid cell sizes.

Table 11: HIV Awareness Index - specification 2 (varied grid cell sizes)

	10km	15km	20km	25km	30km	35km	40km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Arrival *	0.0294	0.0711	0.056	0.048	0.0907	0.061	0.0257
Connected	(0.0692)	(0.0461)	(0.0681)	(0.0666)	(0.0598)	(0.0649)	(0.0664)
Young * Arrival	-0.018	-0.0196	-0.0214	-0.0201	-0.0212	-0.0199	-0.0187
* Connected	(0.0358)	(0.0359)	(0.0348)	(0.0361)	(0.0391)	(0.0366)	(0.0368)
N	20,154	20,154	20,154	20,154	20,154	20,154	20,154
R_2	0.2280	0.2196	0. 2162	0.2129	0.2084	0.2038	0.2050
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid *	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Connected FE							
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* p < 0.10, ** p <0.05, *** p <0.01

Table 12: HIV Prevention Index - specification 3 (varied grid cell sizes)

	10km	15km	20km	25km	30km	35km	40km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Arrival *	-0.0396	-0.128***	0.0034	0.0436	0.0123	0.0004	0.0882
Connected	(0.0726)	(0.0381)	(0.0543)	(0.0687)	(0.0599)	(0.0842)	(0.0696)
Male * Arrival	-0.1424	-0.1382	-0.1335	-0.1518*	-0.1466*	-0.1458	-0.1433
* Connected	(0.1013)	(0.09)	(0.0913)	(0.0898)	(0.0791)	(0.0911)	(0.1057)
N	2,311	2,311	2,311	2,311	2,311	2,311	2,311
R_2	0.1962	0.1901	0.1897	0.1884	0.1886	0.1856	0.1863
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid *	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Connected FE							
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* p < 0.10, ** p <0.05, *** p <0.01

The Protection Index, which has yielded the most conclusive results so far, resembles the Prevention Index with regard to its robustness to grid cell size changes (Table 13). Both the effect of the treatment on the treated as well as that of being male and connected to fast Internet do change slightly in both magnitude and statistical significance. Whereas the treatment effect on the treatment group first increases in statistical significance and then turns insignificant with the rise in grid cell sizes, the effect of being male and connected after the fast Internet arrival varies in significance between the 1% and 5% levels throughout the different grid cell variations. However, the magnitude of coefficients varies primarily for the former coefficient. Thus, the variables of interest seem to be sensitive to grid cell sizes to some extent.

The remaining three indices appear to be robust to grid cell changes when considering the respective specifications (Appendix D, Table 20, 21 and 22). All three indices continue to report insignificant results for all coefficients of interest, indicating that an alternation in grid cell sizes does not offset the lack of a significant effect of fast Internet access on HIV outcomes.

To conclude, sensitivity to grid cell size changes can be observed to some extent for the HIV Prevention and HIV Protection Index which both report some altered results throughout the different grid sizes. However, as statistical significance levels and size of coefficients only vary slightly and do not impose a different interpretation of results, I consider the size change in grid cells to not affect the robustness of both indices to a worrying extent.

Table 13: HIV Protection Index - specification 4 (varied grid cell sizes)

	10km	15km	20km	25km	30km	35km	40km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Arrival *	-0.1317*	-0.1813***	-0.1833***	-0.1046	-0.1071*	-0.0937	-0.109
Connected	(0.0706)	(0.0598)	(0.0578)	(0.0691)	(0.0627)	(0.0651)	(0.0678)
Male * Arrival	0.31**	0.3034**	0.3034***	0.2914**	0.2991***	0.2889**	0.2905***
* Connected	(0.1303)	(0.1275)	(0.1084)	(0.1074)	(0.0988)	(0.1124)	(0.0979)
N	14,317	14,317	14,317	14,317	14,317	14,317	14,317
R_2	0.1175	0.1635	0.1625	0.1606	0.1562	0.1564	0.1550
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.3 Excluding Remote Locations

As a third robustness test, I exclude all individuals from the sample that belong to sampling clusters located more than 5km away from the backbone network. Following Hjort & Poulsen's reasoning, I assume those individuals from the control group that live in more remote locations (i.e. more than 5km away from the network) to differ more substantially from the connected individuals than those living in closer geographic proximity to both, the network and the treatment group. The reason I initially included cluster samples within a 7.5km distance to the network was to balance out the indirect effects of fast Internet access that might be experienced by respondents living very close to the 500m treatment radius. Thus, using a slightly smaller control radius than Hjort & Poulsen (2019), i.e. 7.5km instead of 10km, I sought to find a middle way between accounting for an abundance of indirect effects at a very small radius and the adherence of parallel trends of both groups prior to the treatment decreasing with distance. In the following, I intend to determine whether the baseline estimation is, thus, affected by containing a larger sample including relatively more remote households. Robustness of the model would therefore be provided if the effect of fast Internet access was not substantially changed by this reduction of the control group size.

Results to the adapted specifications are reported in Tables 14, 15 and 16 for HIV Awareness, HIV Prevention and Protection Indices. The remaining indices, Attitude towards HIV, Transmission and the Final Index can be found in Appendix E (Table 23, 24 and 25, respectively). As done in the previous two sub-sections, I once again only display outcomes for the variables measuring the effect of fast Internet access. For all indices - expect for the Protection Index – the adapted sample only results in slight changes in magnitude for the coefficients in each specification, while they all retain their statistical insignificance. Only when considering the HIV Protection Index, one observes statistical significance of being male and being connected to fast Internet access drop from the marginal level to insignificance altogether (specification (3) and (5)). Simultaneously, access to fast Internet increases in significance when measuring the impact on the combined literate treatment group (4): Given the negative sign, literate connected respondents tend to experience a reduction in HIV Protection Awareness of 0.098 (0.1431). Nevertheless, significance levels vary only slightly and do not change the interpretation of results substantially. Consequently, all outcomes can be considered relatively robust to the exclusion of more remotely located respondents. This implies that the baseline estimations did not produce biased results due to including a more diverse control group. However, since the results of the Protection Index were mildly affected by all three robustness tests, related outcomes must be considered with caution.

Table 14: HIV Awareness Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0838 (0.0581)	0.1004 (0.0714)	0.0718 (0.0626)	0.0253 (0.0678)	0.0537 (0.0614)
Young * Arrival * Connected		-0.0304 (0.04)			
Male * Arrival * Connected			0.038 (0.0827)	0.0386 (0.0919)	0.0418 (0.0778)
N	16,756	16,756	16,756	11,566	13,006
R_2	0.1998	0.2039	0.2008	0.2012	0.1748
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Table 15: HIV Prevention Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0918 (0.0847)	-0.0073 (0.1049)	-0.0437 (0.0741)	-0.0756 (0.1151)	-0.0727 (0.0826)
Young * Arrival * Connected		-0.1375 (0.1048)			
Male * Arrival * Connected			-0.1395 (0.1014)	-0.1905 (0.1676)	-0.0865 (0.1239)
N	12,575	12,575	12,575	7,577	8,978
R_2	0.1960	0.1966	0.1966	0.2369	0.1933
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: HIV Protection Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0037 (0.0485)	-0.0219 (0.0551)	-0.0669 (0.0675)	-0.1431** (0.0717)	-0.1002 (0.0734)
Young * Arrival * Connected		0.0216 (0.0595)			
Male * Arrival * Connected			0.1887 (0.1258)	0.2812** (0.1316)	0.1136 (0.1138)
N	17,295	17,295	17,295	11,956	13,322
R_2	0.1489	0.1507	0.1500	0.1569	0.1211
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8. Discussion of Results

8.1 Contribution to Literature

To estimate the impact which the Internet can exert on health behavior and outcomes in developing countries I exploit a novel approach first introduced by Hjort & Poulsen (2019). In order to examine the effect of fast Internet access they take advantage of the arrival of submarine cables at the African coast and the recent compilation of geographic maps of terrestrial networks in African countries. Replicating this method has allowed me to compose a difference-in-difference model by dividing the sample into treatment and control group based on the distance to the closest cable network location. In applying this methodology to the health service sector in a development context I contribute to the literature in three significant ways.

To begin with, I am the first to adopt this method of measuring the effect of fast Internet access by considering the arrival of submarine cables to a different sector. While Hjort & Poulsen (2019) and subsequent authors (Cariolle et al. (2018), Tian (2018) and D'Andrea & Limodio (2019)) all used a similar approach to measure economic impact, e.g. in terms of firm performance, employment or financial technology, I focus on health awareness and behavior.

Second, I investigate the impact of the Internet as a general health information source. In the context of HIV awareness, much focus has been put on testing the effectiveness of individual mass media or social media campaigns, as well as specific Internet-related health programs or targeted websites. The global Internet as a health information provider in itself has been given little attention. This especially holds true with the focus on the African continent where fast and reliable Internet connections are subject to more recent achievements. Thus, compared to previous work, I rather investigate whether individuals which are suddenly connected to fast Internet actually make use of this vast resource to seek out, extract and process information on HIV in order to then apply it in real life.

Third, I am the first one to connect the access to Internet with HIV Awareness and Prevention in the context of Nigeria. While some research has been conducted to investigate the influence of Internet exposure on sexual health and behavior among certain population groups within the country, e.g. the youth (Adebayo et al., 2006; Kunnuji, 2010) or even more specifically, university students (Onasanya et al., 2008), the majority of these studies rather expect online activities to evoke hazardous actions leading to unplanned pregnancies or contraction with STI including HIV/AIDS. Other studies investigate HIV awareness and perception among Nigerians, for instance among high school teachers (Bankole & Mabekoje, 2008), deaf community members (Groce et al., 2007), pregnant women (Okonkwo et al., 2007) and students (Onah et al., 2004). Thus, my study is the first one to interlink both, the use of Internet and HIV Awareness and Prevention while taking into account a representative sample of the Nigerian population as a whole.

8.2 Comparison to Previous Findings

Across all indices I do not find a significant effect of fast Internet access on HIV outcomes. As described above, results for the basic model of specification (1) never reached a desired statistical significance level. When varying the sample groups and restricting respondents based on different characteristics, I only retrieve meaningful results for the Protection Index: access to fast Internet appears to positively affect HIV Protection Awareness among the young, male and literate subgroups. Nevertheless, some of the other indices obtained some insightful partial outcomes, which will all be discussed in the following.

In my study, fast Internet access does not appear to positively change respondents' Attitude towards other people living with HIV. A person's attitude, however, tends to be more challenging to change than knowledge in general. Thus, even when reading facts and objective information, a person might still be consumed by superstition or prejudice, potentially already existent inside that person's subconsciousness since early childhood. Oftentimes attitudes are formed by social interactions and an individual's culture, religion and direct surroundings and can therefore not be easily changed by reading contradicting information. This dilemma has been discussed in several studies. By also using DHS data from Nigeria and the Democratic Republic of Congo, Gebremedhin et al. (2017) found that while HIV/AIDS awareness has risen among young women over years in both countries, attitude towards people living with the disease remained at low rates. Moreover, Thanavanh et al. (2013) found contradicting attitudes persistent in unmarried male high school students in Lao People's Democratic Republic when considering a close relative or friend contracting HIV compared to more distant acquaintances. While they were willing to care for and support a person infected with HIV close to them, they were reluctant to continue visiting shops run by an HIV-infected owner or school classes taught by HIV-positive teachers. Equally, a comparable study in Iran found high school students to hold significant intolerant attitudes towards people with HIV despite relatively high HIV knowledge rates (Tavoosi et al., 2004). Although expressing compassion and empathy towards people suffering from the disease, the students were unwilling to engage in any social interaction including visiting the same school or shaking hands with HIV-infected individuals. The persistence of such negative attitudes implies that subjective misconceptions about HIV might still preserve in spite of increasing knowledge rates. Therefore, these findings indicate that oftentimes attitude does not change alongside increasing awareness and knowledge patterns. Consequently, attitude might not constitute a reliable indicator to measure HIV Awareness and might instead rather be a subsequential objective resulting from awareness over time. Nevertheless, comprehending the risk of HIV/AIDS, its transmission routes and protection measures as well as developing a positive attitude free of stigma towards people with HIV are both crucial to combat infection rates in Africa (Gebremedhin et al., 2017). Thus, while this explains why access to fast Internet alone might not change attitudes among respondents in Nigeria, developing strategies to reduce negative attitude towards the disease should nevertheless turn to the centre of attention to fight HIV prevalence.

Transmission of HIV/AIDS is similarly tainted by sociocultural barriers as people's attitude towards the disease. Consequently, the findings of this study, namely that access to fast Internet does not appear to impose an effect on the Transmission Index, neither for the entire treatment group, nor for any of its restricted subgroups, might stem from identical external factors hindering such change. As discussed above, prevention methods to avoid MTCT are oftentimes considered women's topics, simply ignored by the male members of a household - a social norm also observed in the regression results of specification (3) and (4). Given that I especially consider young men to exploit fast Internet access, the inexistence of a significant increase in HIV Transmission Awareness might thus be explained by these gender-based behavioural expectations in many African communities. A study from Cameroon supports this reasoning by finding traditional gender perceptions and tribal culture to inhibit young men from supporting their wives in MTCT awareness and prevention. Despite knowing that accompanying their wives to training sessions could be effective for the entire family, they retain from doing so in fear of social disapproval (Nkuoh et al., 2010). Adelekan et al. (2014) came to a similar conclusion when interviewing married men in the Nigerian city of Osogbo: Despite knowing about and appreciating the effectiveness of preventive MTCT programs, they refused to accompany their wives due to social norms. Thus, transmission of HIV through MTCT might be a topic too afflicted by social stigma to be actively searched for online by the male target group of this study. Consequently, I expect that for the Internet to effectively advance HIV Awareness in terms of MTCT, offline efforts must be increased beforehand. Only when changing gender-related prejudices and cultural expectations through community work, the Internet might pose a valuable source of transmission-related information in the long-term. That such change is already underway might be apparent given the significant increase in Transmission Awareness among the young sample sub-group (2).

In contrast to the previous two outcomes, Protection Awareness is positively changed by fast Internet access for one specific sub-group: Young men aged 15-34 years old appear to actively use online resources to enhance their knowledge on protection measures. This finding is in line with Oyediran et al. (2011) who determines mass media exposure – next to education levels and economic status – to increase use of protective measures among young, unmarried Nigerian men. It is especially the other two factors – education and economic purchasing power – that explain why it is particularly men that experience increased Protection Awareness once gaining access to the global Internet. The crucial distinction is that access to fast Internet alone is not sufficient: to be able to connect an individual also requires access to an Internet-ready device, i.e. a laptop, or, more likely, a mobile phone. Thus, a person's economic status oftentimes decides whether he or she is able to purchase such a device or visit a cybercafé. Moreover, being educated also increases health literacy, implying that more educated individuals will be more likely to know about diseases and protection methods in the first place and further understand how to search for additional information online. Given that men outperform women in both education and financial means, their increased awareness of HIV protection methods compared to that of women becomes apparent. In addition, there exists a HIV knowledge gap between Nigerian men and women a priori (Lammers et al., 2013). Thus, even if being literate and owning an Internet-ready device, women might not know that protective methods exist to prevent an HIV infection and would therefore not use the Internet to search for additional information. Moreover, Nigerian men are more likely to have engaged in sexual intercourse with multiple partners (Lammers et al., 2013). Consequently, the likelihood of having discussed prevention methods with one of these partners is higher and might have led to subsequent online searches for additional STI information.

The discussion so far undermines the huge variations among the three HIV Awareness sub-indices, which explains for the lack of significant changes in the combined HIV Awareness Index. All three are affected by different external factors including educational levels, social expectations, economic status and gender-related prejudice. The reason why results for the

Protection Index vary from those for the remaining two sub-indices might be the following: gaining knowledge about protection methods is less provocative to previous frames of mind. Instead of actively changing one's personal perceptions towards a certain social behaviour and traditions, rising one's HIV Protection Awareness oftentimes only requires simple information provision – given that sexual protection is a topic the majority of people has heard about in public, also in connection with other STI or simply to avoid pregnancy. Therefore, it can be considered less tainted by prejudice and social stigma. Moreover, it is easier for men to relate to sexual protection than to discussions on topics such as MTCT. In addition, sex protection campaigns promoting the use of condoms (also for fertility and family planning reasons) have been entertained throughout the entire country by a magnitude of different parties and interest groups (e.g. the USAID campaign) – which effectively helped to take down the taboo around this topic at least to a certain degree. Thus, the hurdle to gather additional information might not be too big compared to other HIV-related topics. Reading through simple educative websites might hence be sufficient to enhance Protection Awareness amongst Nigerian Internet users. In this case, the first steps towards an acceptance level have been done offline, whereas when intending to adhere to people's attitude or MTCT knowledge, passive websites might not be appealing. Considering the four different online intervention categories discussed in the literature review, more interactive, long-time interventions might create a higher impact with regard to these more intrinsically rooted topics. However, as these programs tend to be associated with specific offline facility-run health projects, it is not likely that these online interventions are easily accessible on great scale in Nigeria. Hence, online forums, blogs and websites alone might not be sufficient to successfully enhance HIV Attitudes and Transmission Awareness. The lack of a significant effect of fast Internet access on HIV outcomes could thus partly be traced back to the fact that in previous literature the highest impact on health outcomes can be found when applying professional-led interactive interventions and ensuring users' continuity and consistency of usage. This kind of support might simply not constitute the majority of content that Nigerians receive access to over the Internet. HIV Protection Awareness, on the contrary, might have evolved to a socially accepted subject openly discussed to such an extent, that the barrier to searching for additional information online is respectively lower and passive, educational websites suffice.

Access to fast Internet does not appear to significantly enhance HIV Prevention measurements undertaken by the Nigerian population. This finding aligns with the outcomes of a study conducted by Omoyeni et al. (2014): They conclude that an increased knowledge of HIV does not affect active HIV prevention of young Nigerians. Specifically, they found usage of condoms during sexual intercourse to only amount to 1.5% (men) and 1.6% (women) despite knowledge of HIV. The dilemma might be similar to that described in reference to Transmission and Attitude: Despite being aware of the disease and its danger, individuals do not undertake the actions required to prevent an infection. Thus, even though individuals might have the opportunity to look up HIV-related information online, this does not lead to a change in their real-life actions. Lammers et al. (2013) discuss the stigma associated with the disease, which might lead sexual partners to avoid the topic altogether during intercourse. Moreover, even if respondents might use the Internet to gain more knowledge about HIV/AIDS, its prevention methods and treatment measures, they might simply not know how to gain access to HIV tests or condoms. Given the vast geographical area across which the Nigerian population is spread, health facilities and care centers where people can get tested and obtain condoms might not be accessible. Thus, it might be possible that even though the young, male population has gained the knowledge on how to protect themselves from HIV, they lack (the access to) the actual resources to do so. These resources might only be available to those respondents who live in cities contrary to those in more rural areas – given a higher density of medical professionals and support services. Consequently, even in the presence of fast, globally connected Internet, external factors such as social stigma and lack of sufficient health support centers might hinder the actual realization of acquired knowledge.

To conclude, this study does not find a clear relationship between access to fast Internet and HIV Awareness and Prevention in Nigeria. However, it also does not find the contrary effect. Whereas Hong et al. (2006) found that Chinese college students who used the Internet did not only lack knowledge of HIV, but also engaged in more risky sexual behavior triggered through online content, such an impact cannot be found in the case of Nigeria. That HIV prevention methods and Attitude rates are not significantly reduced after the arrival of fast Internet access provides some evidence to this.

8.3 External Validity of the Study

Exactly as the results of the research on the health impact of online interventions in developed countries - which was discussed in the literature review - cannot be extended to the context of Nigeria, generalizing outcomes the other way around is also ineffectual. Public health systems, predominant diseases, treatment opportunities and costs vary substantially between the developed and developing world. Moreover, acceptance and open handling of a disease as well as transmission-related interpretation and prejudice differ a lot, even within countries. Apart from this, wealth indices, education levels and gender norms tend to be significantly higher in developed countries. Consequently, the results obtained from the study apply to the context of Nigeria but cannot be extended to the Western World. To what extent results diverge between both geographic regions can be observed when comparing the outcomes of this study to the research results discussed in the literature review.

Considering other countries in Africa, however, I expect a certain level of generalizability. Countries at the African coastline, which gained access to fast Internet through the arrival of submarine cables between 2009 and 2012 and already had a terrestrial backbone infrastructure in place beforehand, might reach similar results as the ones obtained in this study. However, I expect the prevalence of HIV in that country, as well as the population's public handling of the disease and the associated social stigma to affect the potential of comparability. Moreover, I consider countries with higher literacy rates and gender parity to potentially experience a greater impact of fast Internet access. In addition, the extent to which the national and local governments invest in public HIV awareness campaigns targeted to their populations via the Internet might substantially affect HIV-related outcomes.

Moreover, the matter of time has to be considered in this context. What was state of the art in online interventions and programs in 2013 might already be out of date today. In addition, I expect the range of online offerings hosted by health providers in Nigeria today to depict a multiple of what was available at the time of the last survey. Nigerians might also have more opportunities of actually using the fast Internet access, e.g. due to increased mobile phone possession rates and extended terrestrial infrastructure. Consequently, I withdraw from pronouncing full generalizability of the outcomes to today's setting.

Finally, I do only partly expect the results of this study to be generalizable to the combatting of other diseases through the Internet. Since HIV/AIDS is characterized by prejudicial discrimination and deeply interwoven to a person's sexual activity it represents a taboo subject consciously avoided in many African societies. While these complications can be easily assumed for other STI, I expect awareness and prevention to be more easily changed in people's minds when discussing less socially delicate diseases such as malaria or tuberculosis.

8.4 Implications for Further Research

Within the frame of this study, an effect of fast Internet use on HIV Awareness and Prevention cannot be established and only be scarcely identified for the sub-index measuring HIV Protection. However, this does not imply that such an effect does not exist. To further investigate this subject, I propose to vary the constellation of the indices. As discussed before, the primary indicators defining HIV Prevention available from the DHS survey rounds only amounted to two – whether the individual had ever been tested for HIV and whether the person had used protection during their last sexual intercourse. Thus, the index fails to provide a holistic picture of prevention methods potentially undertaken by respondents. Drawing from other data sets or conducting an own survey specifically designed to create the optimal HIV Awareness and Prevention indices could fill in these gaps.

Moreover, it is important to bear in mind, that the outcomes of this study indicate that access to fast Internet as such does not affect HIV Awareness and Prevention. Thus, effort should now be dedicated to understanding where exactly the barriers lie. Amongst potential reasons are the following: Health illiteracy might hinder Nigerian Internet users to fully grasp what they are reading about HIV threat and protection; most Nigerian Internet users might not know about the existence of health informative online material; Internet users might be too preoccupied by their attitude or opinion on HIV so that they do not believe what they read about the disease online. Consequently, further research could focus on identifying the primary bottlenecks and subsequently establish solutions to overcome them.

Another research path could focus on the sample subgroup of male, literate and young respondents. The outcomes have shown that these subgroups do in general experience significantly better HIV index outcomes compared to older or female respondents. That these groups use the Internet for different purposes as soon as it gets available to them is without doubt: Their Internet usage rates have been sharply rising across the African continent in recent years. Thus, further studies should identify how to create online information content that draws the attention of this target group and steepens their learning curve with regard to health topics. Taking into account the strand of research that has found Internet exposure to enforce a riskier sexual behavior among young, male Nigerians, incentives should be tested that counteract such hazardous development and instead increase HIV Awareness and Prevention.

9. Conclusion

This paper investigates the impact of fast Internet access on HIV Awareness and Prevention in the context of Nigeria. It particularly examines the extent to which the arrival of fast and globally connected Internet affects HIV Attitude, Transmission and Protection Awareness as well as Prevention measures for the specific sample subgroups of young, male and literate Nigerians. Considering the connection of submarine cables to landing points at the African coastline as a technological exogenous shock, I construct a difference-in-difference model to compare the change in HIV outcomes between those individuals subsequently connected to the Internet and those remaining unconnected. The validity of my model is ensured by including individual control characteristics, time-fixed effects as well as location-fixed effects to control for individual as well as temporal- or regional-specific variations. Moreover, three robustness tests are implemented to test whether the model is sensitive to changes in the treatment group size, variations in the grid cell size and the exclusion of remote household locations.

This paper provides the following main findings: First, it finds that access to fast Internet significantly increases HIV Protection Awareness for the male, young as well as literate population. The impact is highest in magnitude for young men, for which the index increases substantially by 0.213 when compared to the overall rather modest rise in Protection Awareness over years. This result implies that the Internet indeed poses an effective opportunity to disseminate HIV-related information targeted to this group specifically. Second, no causal relationship can be identified between fast Internet access and HIV Attitude, Transmission and Prevention. All three factors seem to be rooted so strongly within a person's subconsciousness that information provision alone will not change perceptions or actions. While Attitude is oftentimes biased by superstition or prejudice formed through a person's cultural and religious environment since childhood, Transmission – especially MTCT transmission routes – are afflicted with gender-related expectations. Moreover, Prevention might be limited by both social stigma as well as lack of access to respective resources and health care facilities, thus hindering individuals to take action despite having acquired the necessary HIV-related knowledge. Thus, in all three cases, the Internet as an information provider alone is not sufficient. Designing tailored, self-guided program interventions accompanied by offline counseling instead, might overcome this burden.

While the positive effect of fast Internet access on HIV Protection Awareness is proven, little is known about the exact online channels through which information dissemination is most effective. Consequently, further research should focus on identifying the key online interventions primarily accepted and used by the distinct target groups. Learning more about whether it is informative websites, online forums or interactive self-guided interventions that deliver HIV knowledge most clearly and comprehensibly, will help future policy makers as well as both public and private health care providers to design prospective online protection campaigns in the most efficient manner.

List of References

Abiodun, M. O., Munir'deen, A. I., & Aboyeji, P. A. (2007). Awareness and knowledge of mother-to-child transmission of HIV among pregnant women. *Journal of the National Medical Association*, *99*(7), 758 – 763

Adebayo, D.O., Udegbe, I. B., & Sunmola, A. M. (2006). Gender, Internet use, and sexual behavior orientation among young Nigerians. *CyberPsychology & Behavior*, *9*(6), 742 – 752

Ad Hoc Committee on Health Literacy (1999) Health literacy: report of the Council on Scientific Affairs, American Medical Association. *Journal of the American Medical Association*, 281, 552 – 557

Adelekan, A. L., Edoni, E. R., & Olaleye, O. S. (2014). Married men perceptions and barriers to participation in the prevention of mother-to-child HIV transmission care in Osogbo, Nigeria. *Journal of Sexually Transmitted Diseases*, 680962

Alpay, L., Verhoef, J., Xie, B., Te'eni, D., & Zwetsloot-Schonk, J. H. M. (2009). Current challenge in consumer health informatics: Bridging the gap between access to information and information understanding. *Biomedical Informatics Insights.* 2, 1 – 10

Al-Shorbaji, N. (2012) Empowering people and organizations through information. *Journal of Health Communication*, 17(2), 1 – 4

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481 – 1495

Andersson, G., & Kaldo, V. (2004). Internet-based cognitive behavioral therapy for tinnitus. *Journal of Clinical Psychology*, *60*(2), 171 – 178

Archibong, E. P., Enang, E. E., & Bassey, G. E. (2017). Witchcraft beliefs in diseases causation and health–seeking behaviour in pregnancy of women in Calabar South–Nigeria. *IOSR Journal of Humanities and Social Sciences*, *22*, 24 – 8

Aregbeshola, B. S., & Khan, S. M. (2018). Out-of-pocket payments, catastrophic health expenditure and poverty among households in Nigeria 2010. *International Journal of Health Policy and Management*, *7*(9), 798 – 806

Avert (2019). HIV and AIDS in Nigeria. Retrieved April 20th, 2020 from https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/nigeria

Awofala, A. A., & Ogundele, O. E. (2018). HIV epidemiology in Nigeria. *Saudi Journal of Biological Sciences*, *25*(4), 697 – 703

Balsa, A. I., & Gandelman, N. (2010). The impact of ICT on health promotion: A randomized experiment with diabetic patients. *IDB Working Paper No. IDB -WP -221*

Bamise, O. F., Bamise, C. T., & Adedigba, M. A. (2011). Knowledge of HIV/AIDS among secondary school adolescents in Osun state, Nigeria. *Nigerian Journal of Clinical Practice*, *14*(3), 338 – 344

Bankole, O. M., & Mabekoje, O. O. (2008). Awareness and opinions about HIV/AIDS among secondary school teachers in Ogun State, Nigeria. *Scientific Research and Essay*, *3*(6), 245 – 253

Barak, A. (2007). Emotional support and suicide prevention through the Internet: A field project report. *Computers in Human Behavior*, *23*(2), 971 – 984

Barak, A., & Grohol, J. M. (2011). Current and future trends in internet-supported mental health interventions. *Journal of Technology in Human Services*, *29*(3), 155 – 196

Berlin, M.P., Bonnier, E., & Olofsgård, A. (2017). The donor footprint and gender gaps. *WIDER Working Paper 2007/130*

Bowen, A. M., Horvath, K., & Williams, M. L. (2007). A randomized control trial of Internet-delivered HIV prevention targeting rural MSM. *Health Education Research*, *22*(1), 120 – 127

Broom, A. (2005). Virtually he@lthy: The impact of internet use on disease experience and the doctor-patient relationship. *Qualitative Health Research*, *15*(3), 325 – 345

Bundorf, M. K., Wagner, T. H., Singer, S. J., & Baker, L. C. (2006). Who searches the internet for health information?. *Health Services Research*, *41*(3p1), 819 – 836

Cariolle, J., Le Goff, M., & Santoni, O. (2018). Broadband infrastructure deployment, digital vulnerability, and local firm performance in developing and transition countries. *Ferdi Development Policies Working Paper 195*

Chair, C. & de Lannoy, A. (2018). Youth, deprivation and the Internet in Africa. *Research ICT Africa*. Retrieved May 3rd, 2020 from https://researchictafrica.net/after-access-survey-papers/2018/After_Access:_youth_and_digital_inequality_in_Africa.pdf

Chukwuemeka, I. K., Fatima, M. I., Ovavi, Z. K., & Olukayode, O. (2014). The impact of a HIV prevention of mother to child transmission program in a Nierian early infant diagnosis centre. *Nigerian Medical Journal*, *55*(3), 204

Cline, R. J., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research*, *16*(6), 671 – 692

D'Andrea, A., & Limodio, N. (2019). High-speed Internet, financial technology and banking in Africa. *BAFFI CAREFIN Centre Research Paper 19124*

Daniluk, J. C., & Koert, E. (2015). Fertility awareness online: the efficacy of a fertility education website in increasing knowledge and changing fertility beliefs. *Human Reproduction*, *30*(2), 353 – 363

DHS (2003). *Reproductive maternal and child health in Eastern Europe and Eurasia: A comparative report* – Chapter 10: Knowledge of HIV/AIDS transmission and prevention. Retrieved May 3rd, 2020 from https://dhsprogram.com/publications/publication-od28-other-documents.cfm

Dickerson, F. (1998). Strategies that foster empowerment. *Cognitive and Behavioral Practice*. 5(2), 255 – 275

Dickerson, S., Reinhart, A. M., Feeley, T. H., Bidani, R., Rich, E., Garg, V. K., & Hershey, C. O. (2004). Patient Internet use for health information at three urban primary care clinics. *Journal of the American Medical Informatics Association*, *11*(6), 499 – 504

Dowling, M., & Rickwood, D. (2013). Online counseling and therapy for mental health problems: A systematic review of individual synchronous interventions using chat. *Journal of Technology in Human Services*, *31*(1), 1 – 21

Dutta, U. P., Gupta, H., & Sengupta, P. P. (2019). ICT and health outcome nexus in 30 selected Asian countries: Fresh evidence from panel data analysis. *Technology in Society*, *59*, 101 – 184

Earnshaw, V. A., & Chaudoir, S. R. (2009). From conceptualizing to measuring HIV stigma: A review of HIV stigma mechanism measures. *AIDS and Behavior*, *13*(6), 1160

Eastin, M. S. (2001). Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer-Mediated Communication*, *6*(4), 0-0

Fox, S. and L. Rainie. 2000. *The online health care revolution: How the web helps Americans take better care of themselves*. Washington, DC: The Pew Internet & American Life Project.

Fox, S., & Rainie, L. (2002). E-patients and the online health care revolution. (E-Health). *Physician executive*, *28*(6), 14 – 18

Gabel, L. L., & Pearsol, J. A. (1993). Taking an effective sexual and drug history: A first step in HIV/AIDS prevention. *Journal of Family Practice*, *37*(2), 185 – 188

Gaston, C. M., & Mitchell, G. (2005). Information giving and decision-making in patients with advanced cancer: A systematic review. *Social Science & Medicine*, *61*(10), 2252 – 2264

Gebremedhin, S. A., Wang, Y., & Tesfamariam, E. H. (2017). Predictors of HIV/AIDS knowledge and attitude among young women of Nigeria and Democratic Republic of Congo: Cross-sectional study. *Journal of AIDS Clinical Research*, *8*(3), 1 – 8

Gilmour, J. A. (2007). Reducing disparities in the access and use of Internet health information. A discussion paper. *International Journal of Nursing Studies*, *44*(7), 1270 – 1278

Goold, P. C., Ward, M., & Carlin, E. M. (2003). Can the Internet be used to Improve sexual health awareness in web-wise young people?. *BMJ Sexual & Reproductive Health*, *29*(1), 28 – 30

Groce, N. E., Yousafzai, A. K., & Van der Maas, F. (2007). HIV/AIDS and disability: Differences in HIV/AIDS knowledge between deaf and hearing people in Nigeria. *Disability and Rehabilitation*, *29*(5), 367 – 371

GSMA (2019). Connected Women: The Mobile Gender Gap Report 2019. Retrieved April 22nd, 2020 from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf

Gustafson, D. H., Hawkins, R., Boberg, E., Pingree, S., Serlin, R. E., Graziano, F., & Chan, C. L. (1999). Impact of a patient-centered, computer-based health information/support system. *American Journal of Preventive Medicine*, *16*(1), 1 – 9

Hartzler, A., & Pratt, W. (2011). Managing the personal side of health: how patient expertise differs from the expertise of clinicians. *Journal of Medical Internet Research*, *13*(3), 1 – 8

Healy, A. M. (2008). MedlinePlus go local: Connecting at-risk populations with health care services. From: *Global information inequalities: bridging the information gap*. Oxford: Chando Publishing, 2008, 236 pp.

Hjort, J. & Poulsen, J. (2019), The arrival of fast Internet and employment in Africa, *American Economic Review,* 109(3), 1032 – 79

Holmes-Rovner, M., Stableford, S., Fagerlin, A., Wei, J. T., Dunn, R. L., Ohene-Frempong, J.,Kelly-Blake, K. & Rovner, D. R. (2005). Evidence-based patient choice: A prostate cancer decision aid in plain language. *BMC Medical Informatics and Decision Making*, *5*(1), 16

Hong, Y., Li, X., Mao, R., & Stanton, B. (2006). Internet use among Chinese college students: Implications for sex education and HIV prevention. *CyberPsychology & Behavior*, *10*(2), 161 – 169

ITU (2013). Study on international Internet connectivity in Sub-Saharan Africa. Retrieved April 20th, 2020 from https://www.itu.int/en/ITU-D/Regulatory-Market/Documents/IIC_Africa_Final-en.pdf

Jaganath, D., Gill, H. K., Cohen, A. C., & Young, S. D. (2012). Harnessing online peer education (HOPE): Integrating C-POL and social media to train peer leaders in HIV prevention. *AIDS Care*, *24*(5), 593 – 600

Johnson, A., & Sandford, J. (2005). Written and verbal information versus verbal information only for patients being discharged from acute hospital settings to home: Systematic review. *Health Education Research*, *20*(4), 423 – 429

Kalichman, S. C., Weinhardt, L., Benotsch, E., & Cherry, C. (2002). Closing the digital divide in HIV/AIDS care: Development of a theory-based intervention to increase Internet access. *AIDS Care*, *14*(4), 523 – 537

Kalichman, S. C., Pellowski, J., & Turner, C. (2011). Prevalence of sexually transmitted co-infections in people living with HIV/AIDS: Systematic review with implications for using HIV treatments for prevention. *Sexually Transmitted Infections*, *87*(3), 183 – 190

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*, *6*(1), 123 – 135

Kende, M., & Rose, K. (2015). Promoting local content hosting to develop the internet ecosystem. *Internet Society Report*, 2 – 35

Kim, J. N., & Lee, S. (2014). Communication and cybercoping: Coping with chronic illness through communicative action in online support networks. *Journal of Health Communication*, *19*(7), 775 – 794

Korp, P. (2006). Health on the Internet: implications for health promotion. *Health Education Research*, *21*(1), 78 – 86

Kunnuji, M. O. N. (2010). The internet and adolescent sexuality in Lagos metropolis, Nigeria: Influence of online sexual activities on real life sexual behaviour of adolescents. *African Journal of Reproductive Health,* 16(2), 207 – 218

Lammers, J., van Wijnbergen, S. J., & Willebrands, D. (2013). Condom use, risk perception, and HIV knowledge: A comparison across sexes in Nigeria. *HIV/AIDS*, 5, 283

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, *4*(3), 165 – 224

Lou, C. H., Zhao, Q., Gao, E. S., & Shah, I. H. (2006). Can the Internet be used effectively to provide sex education to young people in China?. *Journal of Adolescent Health*, *39*(5), 720 – 728

Lustria, M. L. A., Cortese, J., Noar, S. M., & Glueckauf, R. L. (2009). Computer-tailored health interventions delivered over the Web: Review and analysis of key components. *Patient Education and Counseling*, *74*(2), 156 – 173

Mahajan, A. P., Sayles, J. N., Patel, V. A., Remien, R. H., Ortiz, D., Szekeres, G., & Coates, T. J. (2008). Stigma in the HIV/AIDS epidemic: A review of the literature and recommendations for the way forward. *AIDS, 22*, 67 – 79

Makri, A. (2019). Bridging the digital divide in health care. *The Lancet Digital Health*, 1(5), 204 – 205

Mahlknecht, G. (2014). Map of Undersea Cables. Retrieved April 10th, 2020 from https://www.cablemap.info/

McKay, H. G., King, D., Eakin, E. G., Seeley, J. R., & Glasgow, R. E. (2001). The diabetes network internet-based physical activity intervention: A randomized pilot study. *Diabetes Care*, *24*(8), 1328 – 1334

McMullan, M. (2006). Patients using the Internet to obtain health information: How this affects the patient–health professional relationship. *Patient Education and Counseling*, *63*(1-2), 24 – 28

McPherson, C. J., Higginson, I. J., & Hearn, J. (2001). Effective methods of giving information in cancer: A systematic literature review of randomized controlled trials. *Journal of Public Health*, *23*(3), 227 – 234

Mesch, G., Mano, R., & Tsamir, J. (2012). Minority status and health information search: A test of the social diversification hypothesis. *Social Science & Medicine*, *75*(5), 854 – 858

Moskowitz, D. A., Melton, D., & Owczarzak, J. (2009). PowerON: The use of instant message counseling and the Internet to facilitate HIV/STD education and prevention. *Patient Education and Counseling*, *77*(1), 20 – 26

Murray, E., Burns, J., See Tai, S., Lai, R. & Nazareth, I. (2005) Interactive Health Communication Applications for people with chronic disease. *The Cochrane Database of Systematic Reviews*, 4 (1), 1 – 80

Neal, D. M., & McKenzie, P. J. (2011). Putting the pieces together: Endometriosis blogs, cognitive authority, and collaborative information behavior. *Journal of the Medical Library Association*, *99*(2), 127

Nelson, A. (2002). Unequal treatment: Confronting racial and ethnic disparities in health care. *Journal of the National Medical Association*, *94*(8), 666

Neter, E., & Brainin, E. (2012). eHealth literacy: Extending the digital divide to the realm of health information. *Journal of Medical Internet Research*, *14*(1), 19

Nkuoh, G. N., Meyer, D. J., Tih, P. M., & Nkfusai, J. (2010). Barriers to Men's Participation in Antenatal and Prevention of Mother-to-Child HIV Transmission Care in Cameroon, Africa. *Journal of Midwifery & Women's Health*, *55*(4), 363 – 369

OECD (2014). International cables, gateways, backhaul and international exchange points. *OECD Digital Economy Papers No. 232*, 1 – 41

Oenema, A., Brug, J., & Lechner, L. (2001). Web-based tailored nutrition education: Results of a randomized controlled trial. *Health Education Research*, *16*(6), 647 – 660

Okonkwo, K. C., Reich, K., Alabi, A. I., Umeike, N., & Nachman, S. A. (2007). An evaluation of awareness: Attitudes and beliefs of pregnant Nigerian women toward voluntary counseling and testing for HIV. *AIDS Patient Care and STDs*, *21*(4), 252 – 260

Omoyeni, S. T., Akinyemi, A. I., & Fatusi, A. (2014). Adolescents and HIV-related behaviour in Nigeria: Does knowledge of HIV/AIDS promote protective sexual behaviour among sexually active adolescents? *African Population Studies*, *27*(2), 331 – 342

Onah, H. E., Mbah, A. U., Chukwuka, J. C., & Ikeme, A. C. (2004). HIV/AIDS awareness and sexual practices among undergraduates in Enugu, Nigeria. *The Nigerian Postgraduate Medical Journal*, *11*(2), 121 – 125

Onasanya, S. A., Nicholas, I. S., & Onasanya, E. O. (2008). Influence of the internet on the sexual health of university undergraduates in Makurdi, Nigeria. *Nigerian Journal of Guidance and Counselling*, *13*(1), 41 – 53

Oyediran K.A., Feyisetan O.I., Akpan T. (2011). Predictors of condom-use among young never-married males in Nigeria. *Journal of Health, Population, and Nutrition.* 29(3); 273 – 285

Peña-Purcell, N. (2008). Hispanics' use of Internet health information: An exploratory study. *Journal of the Medical Library Association: JMLA*, 96(2), 101

Ramirez, J. R., Crano, W. D., Quist, R., Burgoon, M., Alvaro, E. M., & Grandpre, J. (2002). Effects of fatalism and family communication on HIV/AIDS awareness variations in native American and Anglo parents and children. *AIDS Education and Prevention*, *14*(1), 29 – 40

Riegelman, R., & Persily, N. A. (2001). Health information systems and health communications: Narrowband and broadband technologies as core public health competencies. *American Journal of Public Health*, *91*(8), 1179 – 1183

Robinson, P., & Serfaty, M. (2008). Getting better byte by byte: A pilot randomised controlled trial of email therapy for bulimia nervosa and binge eating disorder. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, *16*(2), 84 - 93

Rhodes, S. D. (2004). Hookups or health promotion? An exploratory study of a chat room-based HIV prevention intervention for men who have sex with men. *AIDS Education and Prevention*, *16*(4), 315 – 327

Rubenstein, E. L. (2012). "Things my doctor never told me": Bridging information gaps in an online community. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1 – 10

Santo, A., Laizner, A. M., & Shohet, L. (2005). Exploring the value of audiotapes for health literacy: A systematic review. *Patient Education and Counseling*, *58*(3), 235 – 243

Schumann, R., & Kende, M. (2013). Lifting barriers to Internet development in Africa: Suggestions for improving connectivity. *Analysys Mason Limited,* 9(1)

Shahab, L., & McEwen, A. (2009). Online support for smoking cessation: A systematic review of the literature. *Addiction*, *104*(11), 1792 – 1804

Skinner, H., Biscope, S., & Poland, B. (2003). Quality of internet access: Barrier behind internet use statistics. *Social Science & Medicine*, *57*(5), 875 – 880

Skinner, T. A., & Norman, R. W. (2012). Variables influencing the likelihood of cardiac dysrhythmias during extracorporeal shock wave lithotripsy. *Canadian Urological Association Journal*, *6*(2), 107

Sligo, F. X., & Jameson, A. M. (2000). The knowledge—behavior gap in use of health information. *Journal of the American Society for Information Science*, *51*(9), 858 – 869

Smith, D. T. (2013). African Americans and network disadvantage: Enhancing social capital through participation on social networking sites. *Future Internet*, *5*(1), 56 – 66

Staples, L. H. (1990). Powerful ideas about empowerment. *Administration in Social Wor*k, 14(2), 29 – 42

Stout, P. A., Villegas, J., & Kim, H. (2001). Enhancing learning through use of interactive tools on health-related websites. *Health Education Research*, *16*(6), 721 – 733

Sudha, R. T., Vijay, D. T., & Lakshmi, V. (2005). Awareness, attitudes, and beliefs of the general public towards HIV/AIDS in Hyderabad, a capital city from South India. *Journal of Medical Sciences*. 59(7), 307 – 316

Tate, D. F., Wing, R. R., & Winett, R. A. (2001). Using Internet technology to deliver a behavioral weight loss program. *Journal of the American Medical Association*, *285*(9), 1172 – 1177

Tavoosi, A., Zaferani, A., Enzevaei, A., Tajik, P., & Ahmadinezhad, Z. (2004). Knowledge and attitude towards HIV/AIDS among Iranian students. *BMC Public Health*, *4*(1), 1 – 6

Thanavanh, B., Harun-Or-Rashid, M., Kasuya, H., & Sakamoto, J. (2013). Knowledge, attitudes and practices regarding HIV/AIDS among male high school students in Lao People's Democratic Republic. *Journal of the International AIDS Society*, *16*(1), 1 – 7

Tian, L. (2018). Division of labor and extent of the market: Theory and evidence from Brazil.

Tonsaker, T., Bartlett, G., & Trpkov, C. (2014). Health information on the Internet: Gold mine or minefield?. *Canadian Family Physician*, *60*(5), 407 – 408

The Economist (2017). Beefing up mobile-phone Internet penetration in Africa: Without connectivity nothing moves. Retrieved April 15th, 2020 from https://www.economist.com/special-report/2017/11/10/beefing-up-mobile-phone-and-internet-penetration-in-africa

The Girl Effect (2016). Girl Effect: State of the Girl Report 2016. Retrieved April 20th, 2020 from https://www.girleffect.org/documents/39/State_Of_The_Girl_Report__Nigeria.pdf

UNAIDS (2020). Nigeria. Retrieved April 28th, 2020 from https://www.unaids.org/en/regionscountries/countries/nigeria

Van der Maas, F., & Otte, W. M. (2009). Evaluation of HIV/AIDS secondary school peer education in rural Nigeria. *Health Education Research*, *24*(4), 547 – 557

Wall, A. F. (2007). Evaluating a health education website: The case of AlcoholEdu. *Journal of Student Affairs, Research and Practice*, *44*(4), 692 – 714

Welcome, M. O. (2011). The Nigerian health care system: Need for integrating adequate medical intelligence and surveillance systems. *Journal of Pharmacy & Bioallied Sciences*, *3*(4), 470

Weller, D. & Woodcock, B. (2013). Internet traffic exchange: Market developments and policy challenges. *OECD Digital Economy Papers No. 207,* 2 – 99

Willis, E., & Royne, M. B. (2017). Online health communities and chronic disease self-management. *Health Communication*, *32*(3), 269 – 278

Winzelberg, A. J., Classen, C., Alpers, G. W., Roberts, H., Koopman, C., Adams, R. E., Ernst, H., Dev, P. & Taylor, C. B. (2003). Evaluation of an internet support group for women with primary breast cancer. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *97*(5), 1164 – 1173

WHO (2015). Global Health Workforce Alliance. Nigeria. Retrieved May 2nd, 2020 from https://www.who.int/workforcealliance/countries/nga/en/

WHO (2019). HIV/AIDS. Retrieved April, 22nd, 2020 from https://www.afro.who.int/health-topics/hivaids

World Bank (2019). World Bank Open Data. Retrieved April 20th, 2020 from https://data.worldbank.org/

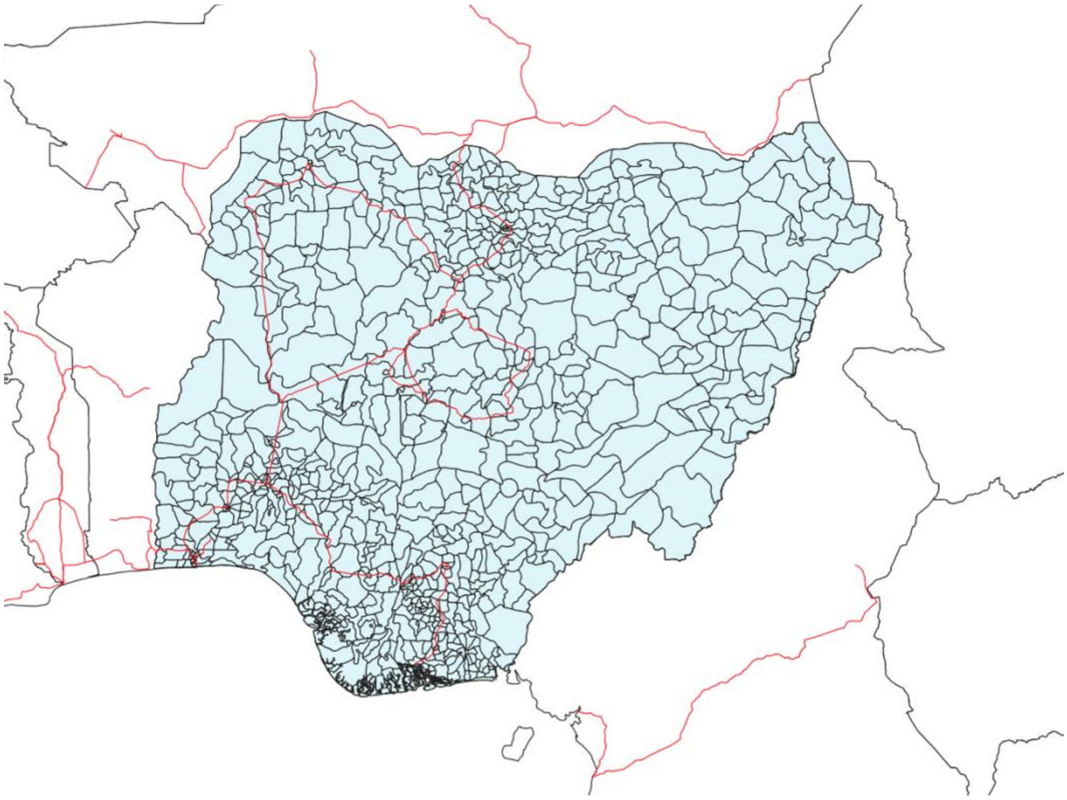
World Bank (2019). The World Bank in Nigeria. Retrieved April 20th, 2020 from https://www.worldbank.org/en/country/nigeria/overview

Ybarra, M. L., & Eaton, W. W. (2005). Internet-based mental health interventions. *Mental Health Services Research*, *7*(2), 75 – 87

Appendix

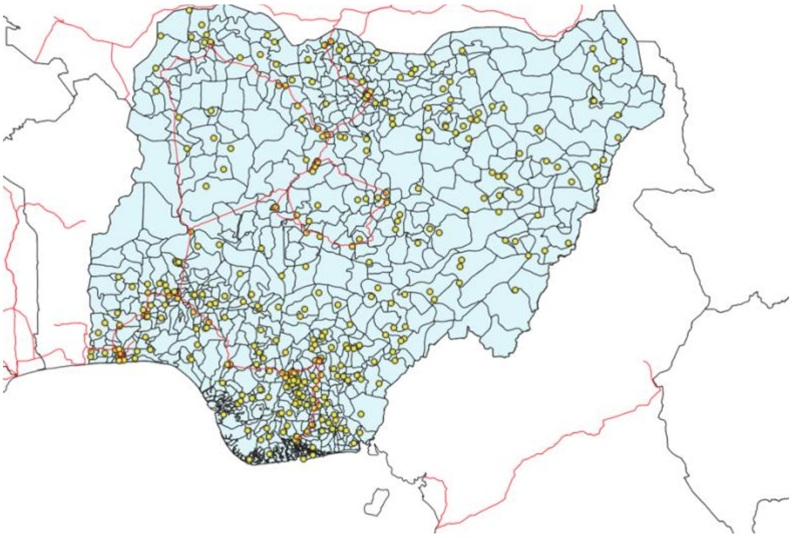
Appendix A: Internet Infrastructure Nigeria

Figure 2: Terrestrial network Nigeria (status 2009)



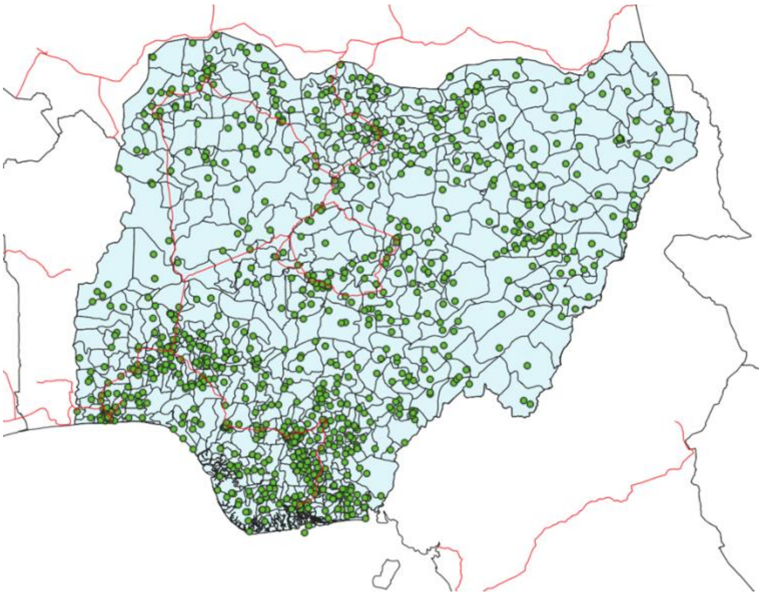
The map displays the territory of Nigeria. The terrestrial backbone infrastructure (red lines) is depicted as it looked like in 2009, the year before the first submarine cable arrived at the Nigerian coast.

Figure 3: 1st survey round Nigeria (2003)



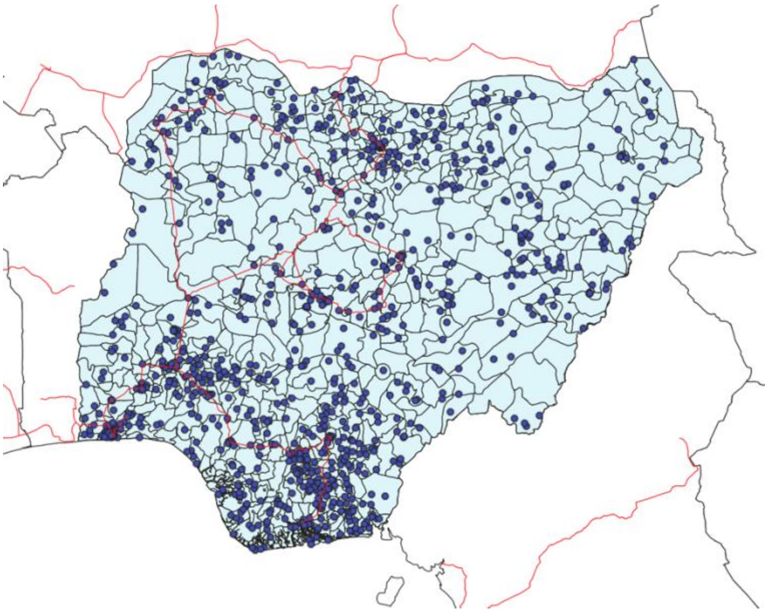
The map displays the territory of Nigeria. The terrestrial backbone infrastructure (red lines) is depicted as it looked like in 2009, the year before the first submarine cable arrived at the Nigerian coast. The sampling clusters of the 2003 survey respondents are presented as the yellow dots.

Figure 4: 2nd survey round Nigeria (2008)



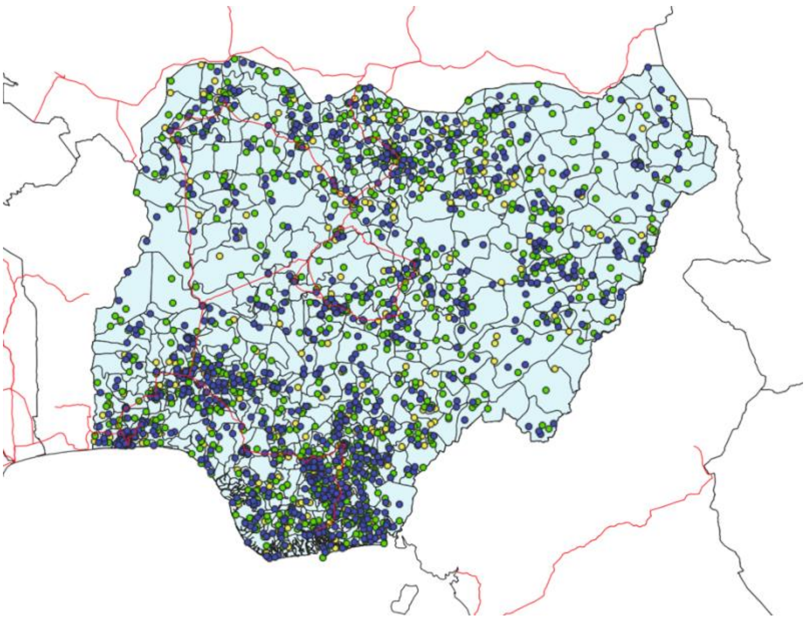
The map displays the territory of Nigeria. The terrestrial backbone infrastructure (red lines) is depicted as it looked like in 2009, the year before the first submarine cable arrived at the Nigerian coast. The sampling clusters of the 2008 survey respondents are presented as the green dots.

Figure 5: 3rd survey round Nigeria (2013)



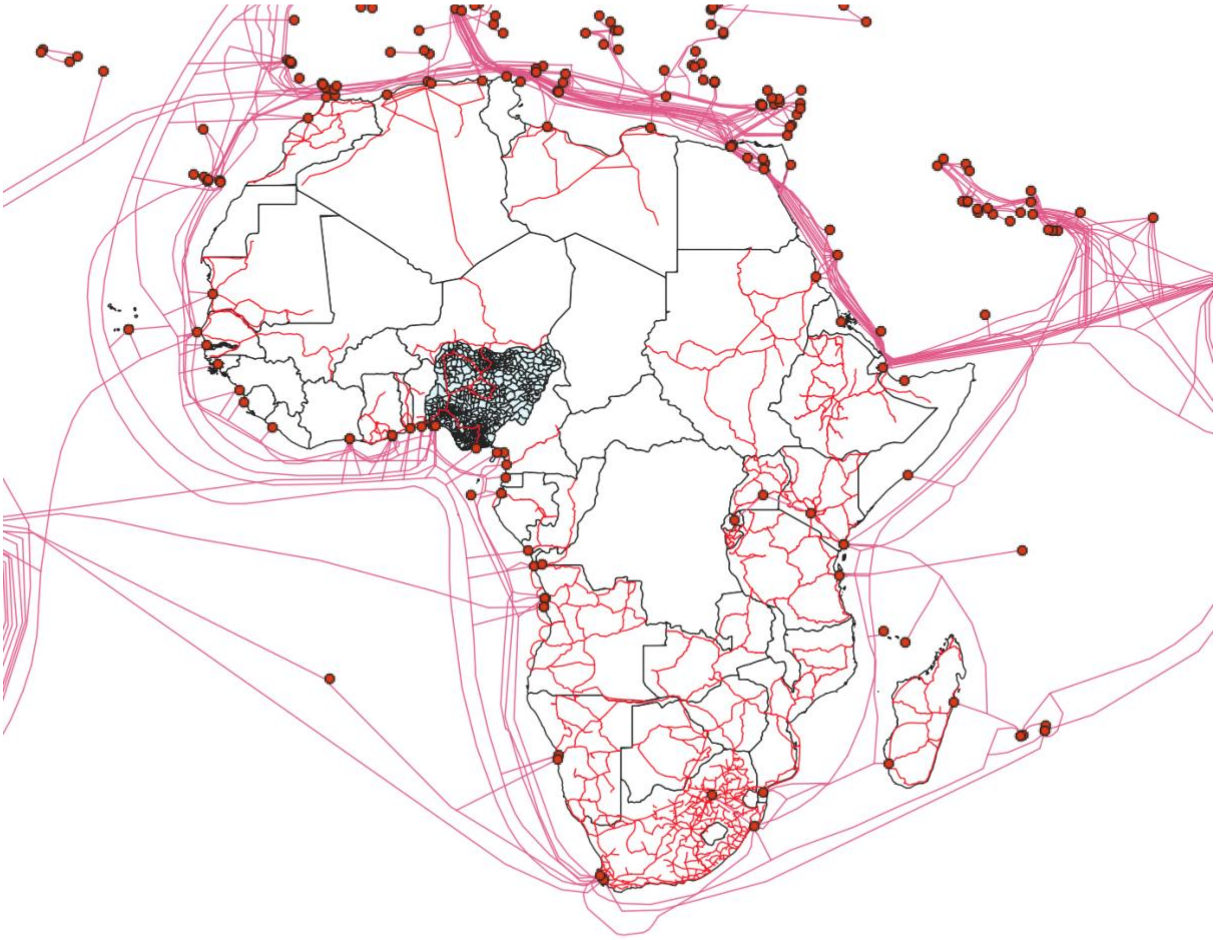
The map displays the territory of Nigeria. The terrestrial backbone infrastructure (red lines) is depicted as it looked like in 2009, the year before the first submarine cable arrived at the Nigerian coast. The sampling clusters of the 2013 survey respondents are presented as the blue dots.

Figure 6: All survey rounds combined (2003, 2008, 2013)



The map displays the territory of Nigeria. The terrestrial backbone infrastructure (red lines) is depicted as it looked like in 2009, the year before the first submarine cable arrived at the Nigerian coast. The sampling clusters of all survey respondents are presented as follows: yellow (2003), green (2008) and blue (2013).

Figure 7: Submarine cables, landing points and terrestrial backbone networks in Africa



I am using Mahlknecht's map of submarine cables to show the submarine cables connecting Africa with the rest of the world as well as the landing points at the coastline (red dots). Given the map was last updated in 2014, it shows more submarine cables arriving at the Nigerian coastline, additional to the one considered in this study: the MainOne cable, which was connected to the Nigerian backbone network in 2010. The terrestrial backbone network (status 2009) is again depicted in red.

Appendix B: Overall Index Trends (2003 – 2013)

Figure 8: Attitude towards HIV Index - overall time trend

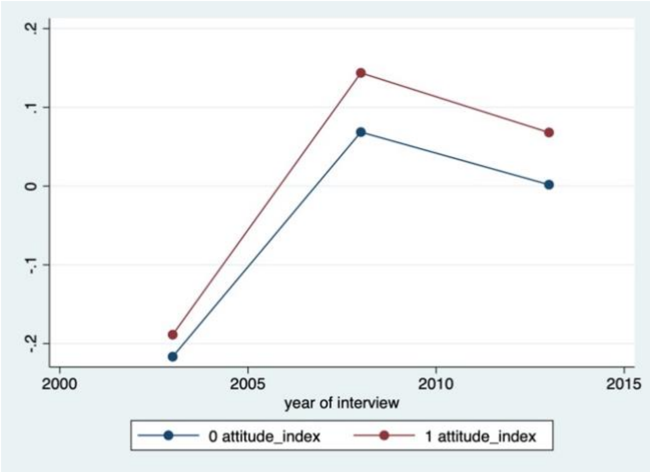


Figure 9: HIV Protection Index - overall time trend

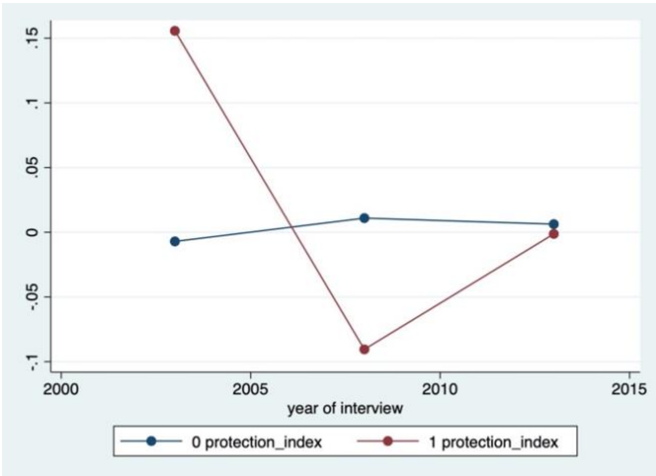
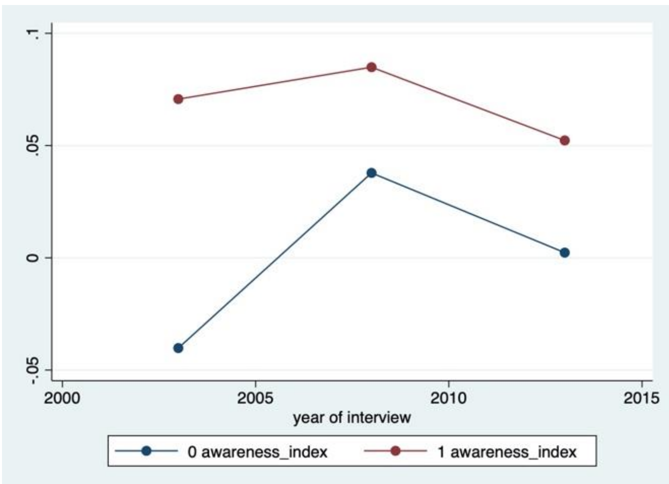


Figure 10: HIV Awareness Index - overall time trend



Appendix C: Varying the Treatment Distances (Robustness Test 1)

Table 17: Attitude towards HIV Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0574 (0.064)	0.058 (0.0635)	0.0802 (0.0683)	0.0448 (0.0784)	0.0608 (0.0709)
Young * Arrival * Connected		-0.0032 (0.0397)			
Male * Arrival * Connected			-0.0698 (0.0804)	-0.0257 (0.0891)	-0.022 (0.0814)
N	21,383	21383	21383	14,727	15,870
R_2	0.1749	0.1757	0.1752	0.1713	0.1551
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Table 18: HIV Transmission Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.0384 (0.0473)	0.0157 (0.0546)	-0.0619 (0.0592)	-0.103 (0.0644)	-0.026 (0.0655)
Young * Arrival * Connected		-0.0814* (0.0459)			
Male * Arrival * Connected			0.0741 (0.0813)	0.0822 (0.0928)	0.0641 (0.077)
N	20,557	20,557	20,557	14,113	15,293
R_2	0.1166	0.1171	0.1168	0.1156	0.1025
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* p < 0.10, ** p <0.05, *** p <0.01

Table 19: Final Index (1km distance threshold)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.1028 (0.267)	0.0829 (0.2498)	0.0208 (0.2896)	0.0934 (0.3468)	0.0055 (0.3217)
Young * Arrival * Connected		0.0175 (0.1951)			
Male * Arrival * Connected			0.2616 (0.3041)	0.0056 (0.3837)	0.0721 (0.3003)
N	13,981	13,981	13,981	8,460	9,719
R_2	0.0763	0.0776	0.0765	0.0894	0.0727
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D: Varying the Grid Cell Sizes (Robustness Test 2)

Table 20: Attitude towards HIV Index - specification 1 (varied grid cell sizes)

	10km	15km	20km	25km	30km	35km	40km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Arrival *	-0.0156	0.0316	0.0299	0.0928	0.0887	0.0651	-
Connected	(0.0897)	(0.095)	(0.0932)	(0.0729)	(0.0745)	(0.0662)	0.0014
							(0.079)
N	21383	21,383	21,383	21,383	21,383	21,383	21,383
R^2	0.1728	0.1584	0.1552	0.1473	0.1563	0.1423	0.1428
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21: HIV Transmission Index - specification 3 (varied grid cell sizes)

	10km	15km	20km	25km	30km	35km	40km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Arrival *	-0.0469	-0.0885	-0.0479	-0.0515	-0.0021	-0.0606	-0.0127
Connected	(0.0728)	(0.096)	(0.0674)	(0.0725)	(0.0658)	(0.0755)	(0.0717)
Male *Arrival	0.0437	0.0446	0.049	0.0431	0.047	0.0577	0.0525
* Connected	(0.0923)	(0.0965)	(0.0951)	(0.1007)	(0.1006)	(0.0993)	(0.1002)
N	20,557	20,557	20,557	20,557	20,557	20,557	20,557
R^2	0.1166	0.1067	0.1007	0.0977	0.0979	0.0921	0.0920
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Final Index - specification 5 (varied grid cell sizes)

	10km (1)	15km (2)	20km (3)	25km (4)	30km (5)	35km (6)	40km (7)
Arrival x Connected	-0.3998 (0.6516)	-0.64 (0.4092)	-0.261 (0.5151)	0.151 (0.5814)	-0.0257 (0.5233)	-0.2191 (0.4745)	-0.1163 (0.497)
Male x Connected x Arrival	-0.0648 (0.4153)	0.0188 (0.4104)	-0.0346 (0.4297)	-0.0185 (0.4703)	-0.0785 (0.4736)	-0.04 (0.4449)	-0.0588 (0.4537)
N	9,719	9,719	9,719	9,719	9,719	9,719	9,719
R_2	0.0672	0.0637	0.0584	0.0534	0.0515	0.0480	0.0528
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects, using the same grid-cell size as stated in the headline.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E: Excluding Remote Locations (Robustness Test 3)

Table 23: Attitude towards HIV Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0109 (0.0949)	-0.0136 (0.084)	0.0177 (0.1085)	0.0012 (0.1342)	0.0694 (0.1186)
Young * Arrival * Connected		0.0327 (0.061)			
Male * Arrival * Connected			-0.0212 (0.1078)	-0.0102 (0.1196)	0.0091 (0.1076)
N	17,827	17,827	17,827	12,323	13,773
R^2	0.1532	0.1540	0.1534	0.1522	0.1480
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 24: HIV Transmission Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	0.0036 (0.0669)	0.0591 (0.1035)	0.0053 (0.0803)	-0.0295 (0.071)	0.0615 (0.0802)
Young * Arrival * Connected		-0.0831 (0.0775)			
Male * Arrival * Connected			-0.0061 (0.1008)	-0.0266 (0.1121)	-0.028 (0.1036)
N	17,096	17,096	17,096	11,783	13,256
R^2	0.1121	0.1125	0.1133	0.1113	0.1047
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 25: Final Index (excluding remote locations)

	Specification (1)	Specification (2)	Specification (3)	Specification (4)	Specification (5)
Arrival * Connected	-0.167 (0.122)	-0.062 (0.1128)	-0.1213 (0.1157)	-0.265* (0.1527)	-0.1775 (0.1281)
Young * Arrival * Connected		-0.1639 (0.1054)			
Male * Arrival * Connected			-0.1239 (0.1063)	-0.1956 (0.1608)	-0.0636 (0.1227)
N	11,397	11,397	11,397	6,860	8,344
R_2	0.0973	0.0999	0.1035	0.1260	0.1033
Time FE	Yes	Yes	Yes	Yes	Yes
Grid * Connected FE	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes

Robust standard errors (in parentheses) are clustered at the level of location fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$