Durham University Business School

Department of Economics

More Information, Less Knowledge: Evidence from Natural and Agentic General Intelligence Experiments

Ben Toas, Muhammad Habibur Rahman & Bibhas Saha

Working Paper No. 17, 2026

Durham Business School working papers series ISSN 1749-3641

More Information, Less Knowledge: Evidence from Natural and Agentic General Intelligence Experiments

Ben Toas*[a] Muhammad Habibur Rahman[a] Bibhas Saha[a]

[a]Durham University

Monday 9th November, 2026

Abstract

Internet access lowers the cost of information search but also changes the content and sources competing for attention. We develop a conceptual framework in which digital access can reduce verified knowledge when a larger information set raises processing costs or displaces attention from higher-quality sources. The first design is a novel 'Agentic General Intelligence Experiment' that randomly assigns survey-profiled language-model agents to controlled information environments. High-quality media feeds raise HIV knowledge more than noisy online environments. Replacing traditional information with a noisy online feed decreases maternal HIV knowledge by 10.7 percentage points. The second design links the staggered expansion of fibre-optic infrastructure in six West African countries to Demographic and Health Survey data. Fibre proximity lowers women's knowledge of HIV transmission during pregnancy by 0.099 points on a 0–2 scale, and a four-item maternal HIV knowledge index by 6.7 percentage points. Internet and social-media use rise, as do "don't know" responses. Exposure to family-planning messages through radio and newspapers and contact with health workers fall.

JEL Codes: D83, I15, O33

Keywords: Internet Access, Information, Health, Artificial Intelligence

*Corresponding author: ben.toas@durham.ac.uk. Any errors in this manuscript are my own. I have no material financial interests related to the research described in this paper.

1 Introduction

Internet access has removed many of the physical constraints on acquiring information. Constraints remain, however, because attention is scarce, source quality varies, and users must decide which claims deserve weight. Online search and social media dilute official guidance with personal opinion and incomplete summaries of correct information. The resulting information set may contain more useful material in absolute terms, compared to a disconnected environment, and still make the verified answer harder to identify. Health knowledge provides a demanding case because small errors can affect testing when symptoms are dismissed and treatment when medical scepticism rises.

This paper studies how digital access changes verified knowledge. We establish a conceptual framework describing the formation of knowledge in a digital information environment. The framework captures both the potential benefits and costs of digital access. On one hand, greater access to information may increase knowledge by expanding the quantity of available information. On the other, digital environments may reduce knowledge through information overload, where the volume of information exceeds individuals' ability to effectively process it. These effects depend critically on the quality of available information.

The framework also distinguishes objective information quality from the weight individuals place on it. Digital connectivity can lower government approval (Guriev et al., 2021), and political trust (Lorenz-Spreen et al., 2023). Reduced institutional trust can lower the weight given to its health messages even if the content is accurate. Competition for attention and algorithmic curation can change the content users encounter (Berger and Milkman, 2012; Chen and Suen, 2023; Robertson et al., 2023; Levy, 2021; Acemoglu et al., 2024). The framework further allows digital media to displace radio, newspapers, and other established sources because attention is scarce. These forces generate an ambiguous prediction for verified knowledge, making the net effect of digital access an empirical question.

We test the model implications in two ways: a randomised Agentic General Intelligence Experiment (AGIE) conducted on survey-profiled agents in a local large-language model (LLM), and a natural experiment from the rollout of fibre-optic infrastructure in West Africa. The AGIE assigns different information feeds to agents and uses the same maternal HIV questions asked in the Demographic and Health Surveys (DHS). The natural experiment estimates the response of maternal HIV knowledge when local access conditions change.

We then examine the effect of expanded internet access using the annual development of terrestrial fibre routes and network nodes. We link these maps to repeated Demographic and Health Surveys (DHS) in Burkina Faso, Ghana, Guinea, Mali, Senegal, and Sierra Leone between 2003 and 2023. The preferred DHS estimates use the road-restricted sample defined below. Respondents are assigned to 10-kilometre by 10-kilometre grid cells. Grid-cell fixed effects absorb time-invariant spatial heterogeneity, while region-by-survey-year

fixed effects absorb shocks common to respondents in the same region and year. The treatment records proximity to infrastructure that supports fixed and mobile broadband. Internet-use data from Afrobarometer confirm that the mapped treatment raises reported internet use by 0.191 points on a five-category frequency scale.

Maternal HIV knowledge falls following fibre exposure. Among women, knowledge that HIV can pass from mother to child during pregnancy declines by 0.099 points on the 0–2 knowledge scale. Knowledge of transmission during delivery and breastfeeding falls by 0.107 and 0.080 points, respectively, and the four-item maternal index falls by 6.7 percentage points. The pregnancy estimate remains between −0.100 and −0.099 as controls are added. The decline therefore spans the measured routes of mother-to-child transmission and is stable across the benchmark specifications.

Three sets of outcomes describe the accompanying change in the information environment. First, fibre proximity raises internet and social-media news use and increases "don't know" responses to each maternal HIV question by 3.0–3.7 percentage points. Second, reported exposure to family-planning information through radio and newspapers falls, as does contact with health workers. Both television-source and antenatal-visit estimates are small and statistically insignificant. General radio news remains stable and television news rises, so the family-planning measures capture changes in exposure to health content rather than a uniform decline in traditional-media use. Third, trust declines across public institutions and public media, while this response is not seen in private media settings. The combined evidence links greater digital exposure to more uncertainty and less contact with some repeated, identifiable health sources.

Subgroup estimates are most negative among women who can read part of a sentence: their implied treatment effect is −0.141 points, compared with −0.075 among women unable to read and −0.118 among women able to read a full sentence. Point estimates are also more negative in households with television access and in the upper three wealth quintiles. These groups are seemingly more likely to access the internet following nearby infrastructure upgrades. The literacy pattern is consistent with a screening effect: partial literacy expands access to written material without necessarily providing the evaluative skills needed to distinguish reliable medical guidance from misleading claims. This interpretation is suggestive because the surveys do not observe the information each respondent encountered or the weight she assigned to it.

The design faces the familiar problems created by non-random infrastructure placement and staggered treatment timing. Fibre routes tend to follow roads, cities, and commercial corridors, so the road-restricted sample is the DHS benchmark. For staggered timing, the preferred modern difference-in-differences specification stacks treatment cohorts and prevents already-treated cells from serving as controls (Wing et al., 2024). Its estimates are −0.074 for pregnancy knowledge and −0.059 for the maternal index; both are statistically significant at the 1 per cent level. Additional checks use an unrestricted comparison sample, vary fixed effects and clustering, exclude major cities and one country at a time, change

the spatial treatment definition, and examine placebo outcomes. Their full results appear in the appendix.

The AGIE uses the characteristics of Senegalese women observed in the DHS to construct a fixed population of language-model agents. Each agent is assessed on seven HIV-related questions: three exactly emulate the maternal HIV questions in DHS and four are wider knowledge checks.[1] Each agent participates in five treatment scenarios evaluated independently, with no transfer of context or memory across scenarios. The feeds vary the shares of accurate, confusing, misleading, and irrelevant content. The final scenario tests the effects of switching from a high-quality feed to a noisy online-like feed.[2] Random assignment identifies differences across these information environments while holding the model, prompt, and survey profiles fixed.

Concise traditional media messages, mimicking health information via the radio or newspaper, raise knowledge by 18–20 percentage points relative to receiving no new information. These gains become smaller as content quality falls, reaching 7.8 percentage points for the three-question maternal HIV score under a noisy online feed. The noisy online feed reduces the seven-question score by 2.3 percentage points. Replacing a high-quality traditional feed with the noisy online feed reduces the three-question score by 10.7 percentage points. The experiment therefore distinguishes the effect of receiving new information from the effect of replacing an established source.

The paper contributes to the broad literature on the economic and political effects of digital infrastructure. The arrival of fast internet increased employment in African locations connected to terrestrial fibre networks (Hjort and Poulsen, 2019). The spread of mobile internet has also affected electoral participation and competition in South Africa (Donati, 2023), while cross-country evidence shows that 3G coverage can reduce confidence in government by making corruption more visible (Guriev et al., 2021). Whereas this work focuses on labour-market and political outcomes, we examine a form of human capital for which accuracy is itself the outcome: whether respondents retain verified health facts.

The closest work on digital access and health addresses different outcomes or relies on different measures of access. Amaral-Garcia et al. (2022) study how online information changes maternal demand and patient choice in England, finding that broadband diffusion increased caesarean sections without improving maternal or infant health outcomes. In Africa, Mensah et al. (2022) link mobile-network expansion to lower infant mortality, while Kergall and Guiffard (2024) examine bed-net use, antenatal care, and childhood immunisation following broadband arrival in Senegal, defining exposure only through proximity to backbone cables. Other studies relate self-reported internet access or ownership of mobile phones to HIV knowledge and reproductive-health behaviour (Iyamu et al., 2021; Iacoella et al., 2022b). Iacoella et al. (2022a) study HIV-related stigma in

[1]This includes: Is there a cure for HIV? Can mosquitos transmit HIV? Can you get HIV from kissing? Can a healthy-looking person have HIV?

[2]The model itself cannot connect to the internet, the feed represents how an online information feed might look like with social media posts, comments and news headlines. Therefore, the reference to an 'online' feed represents an online-like environment.

Ghana, treating improved HIV knowledge as a potential channel. A closely related paper is Iacoella and Tirivayi (2024), who combine DHS data with mobile-phone tower locations across 29 African countries and find that proximity to broadband towers is associated with higher HIV testing awareness. In contrast, using fibre routes and network nodes to measure exposure to fast internet, we find that internet expansion reduces women's knowledge of specific facts about mother-to-child HIV transmission in six West African countries. This may at first appear contradictory, but an increase in confusion on specific causes of transmission, arising from digital media, may indeed motivate individuals to seek out testing facilities.

A second contribution concerns the production and processing of information. Competition for attention affects the accuracy, salience, and completeness of messages (Bordalo et al., 2016; Galperti and Trevino, 2020; Chen and Suen, 2023). Search and rational-inattention models explain why receivers stop before reviewing every source or give more weight to easier material (Caplin et al., 2011; Dean and Neligh, 2023; Guan et al., 2023). Serra-Garcia (2026) provides experimental evidence that attention incentives reduce the information contained in summaries: some readers are induced to seek the full article, but the larger group that stops at the summary becomes less informed. Our evidence connects these ideas to a large change in digital access and to measured health knowledge.

The Agentic General Intelligence Experiment provides the paper's third contribution. Randomised prompts offer a low-cost environment for studying content that would be difficult or unethical to assign to people. The estimand remains internal to the model, prompt, profiles, and message library. We treat the method as a complement to human data and set out the reporting, replication, and validation requirements needed for later applications.

The paper proceeds as follows. Section 2 contains the theoretical model. Section 3 defines the AGIE and reports its estimates. Section 4 presents the fibre-rollout setting, empirical design, and results. Section 5 concludes. Additional natural-experiment and agentic results are collected in the appendix.

2 Conceptual Framework

2.1 Overview

Human capital has long been treated as both a productive input and an accumulable stock. In the Uzawa–Lucas framework, individuals divide scarce time between current production and the acquisition of human capital, which subsequently raises output (Uzawa, 1965; Lucas, 1988). At the individual level, Ben-Porath (1967) models human capital as the output of a production process that combines time, existing skills, and other inputs. Grossman (1972) applies the same stock-flow logic to health: the current stock depreciates but can be augmented through investment. We adapt this framework to health knowledge.

The relevant investment input is exposure to information, and the return to that exposure depends on the source, the quality of its content, the weight an individual places on it, and the individual's ability to process it.

Internet access changes this production process in many ways. It expands the quantity of available information, changes the sources competing for attention, and may alter both the quality of content and trust in its source. These changes do not imply that internet access must always reduce knowledge. Access raises the next-period knowledge stock when it directs attention toward sufficiently valuable information or improves the individual's ability to find and process it. It lowers the next-period stock when substitution toward less useful sources or the cost of processing additional content dominates.

2.2 Information-based human capital

Consider an individual i who enters period t with a stock of health knowledge H_{it} and allocates one unit of attention across information sources indexed by $j = 1, \ldots, N$. Let $a_{ijt} \geq 0$ denote the share of individual i's attention allocated to source j, subject to

$$\sum_{j=1}^{N} a_{ijt} = 1. \tag{1}$$

The constraint captures the opportunity cost of consuming information from competing sources. It follows the central insight of Simon (1971): an abundance of information makes attention, rather than information itself, scarce. Rational-inattention models formalize this constraint by making the acquisition and processing of information costly (Sims, 2003; Matějka and McKay, 2015). Experimental evidence also shows that individuals adjust attention in response to incentives, such as clickbait or sensationalist content, though this feedback loop is not formalised in the framework.

Let $\bar{V}_{jt} \geq 0$ denote the quantity of information supplied by source j per unit of attention. This quantity is taken as given by the individual. The amount of content processed by individual i from source j is therefore $a_{ijt}\bar{V}_{jt}$. Sources differ both in volume and in effective quality. Define

$$Q_{ijt} = \delta_{ijt} q_{jt}, \tag{2}$$

where $q_{jt} \in [-1, 1]$ is the objective quality of information supplied by source j and $\delta_{ijt} \in [0, 1]$ is individual i's trust in that source. Accurate and informative content has $q_{jt} > 0$, irrelevant information has $q_{jt} = 0$, and content that induces incorrect beliefs has $q_{jt} < 0$. Trust determines how strongly the individual incorporates that content. Greater trust improves learning from an accurate source but increases the damage caused by misleading information; scepticism has the opposite effects.

Define individual i's effective information and total information exposure as

$$I_{it} = \sum_{j=1}^{N} a_{ijt} \bar{V}_{jt} Q_{ijt}, \qquad \Omega_{it} = \sum_{j=1}^{N} a_{ijt} \bar{V}_{jt}. \tag{3}$$

Information quality determines how exposure changes knowledge, so it enters I_{it}. Cognitive burden instead depends on the amount of content the individual attempts to process. Accurate, inaccurate, and irrelevant claims all consume processing capacity, so Ω_{it} measures raw exposure without quality adjustment.

Let $A_i > 0$ denote individual i's ability to convert effective information into knowledge. The parameter can depend on predetermined characteristics such as education, literacy, age, and social networks. Health knowledge follows the law of motion

$$H_{i,t+1} = \max\left\{0,\ (1-\rho)H_{it} + A_i I_{it} - \lambda_i \Omega_{it}^2\right\}, \qquad \rho \in [0,1], \quad \lambda_i \geq 0. \tag{4}$$

The initial stock is non-negative, and the zero floor preserves this restriction in subsequent periods. The parameter ρ is the depreciation rate of existing knowledge. The term $A_i I_{it}$ captures learning from newly processed information, while $\lambda_i \Omega_{it}^2$ captures the increasing cognitive cost of processing a larger information set. The quadratic term is a representation of information overload: the marginal contribution of additional information falls as exposure rises and may become negative once information exceeds the receiver's processing capacity (Jacoby et al., 1974; Eppler and Mengis, 2004; Abeler et al., 2025). The specification also accords with economic models in which attention and information processing are costly (Sims, 2003; Dean and Neligh, 2023).

Subtracting H_{it} from both sides of equation (4) gives

$$\Delta H_{it} = \max\left\{-H_{it},\ A_i I_{it} - \lambda_i \Omega_{it}^2 - \rho H_{it}\right\}. \tag{5}$$

Here, $\Delta H_{it} \equiv H_{i,t+1} - H_{it}$. The stock H_{it} is inherited at the start of period t and is therefore predetermined with respect to changes in access during that period. For $H_{it} > 0$, equation (5) allows the knowledge stock to decline: $H_{i,t+1} < H_{it}$ whenever $A_i I_{it} - \lambda_i \Omega_{it}^2 < \rho H_{it}$.

2.3 A two-input model: traditional and modern information sources

For the empirical application, we group sources into traditional media, T, and modern or internet-enabled media, M. Traditional sources include radio, newspapers, state television, and in-person professional advice. Modern sources include online search, social media,

streaming services, and other internet-based channels. Television could potentially be ambiguous as it carries conventional state and private broadcasts, but internet-connected televisions can also provide access to streaming services and platforms such as YouTube.[3]

Let $a_{it} \equiv a_{iMt}$, so $a_{iTt} = 1 - a_{it}$. To simplify notation, we suppress the individual subscript from equation (6) onward, writing a_t for a_{it}.[4] When $H_{t+1} > 0$, equation (5) becomes

$$\begin{aligned} \Delta H_t = A\left[(1-a_t)\bar{V}_{Tt}Q_{Tt} + a_t\bar{V}_{Mt}Q_{Mt}\right] \\ - \lambda\left[(1-a_t)\bar{V}_{Tt} + a_t\bar{V}_{Mt}\right]^2 - \rho H_t, \end{aligned} \tag{6}$$

where $Q_{st} = \delta_{st}q_{st}$ for $s \in \{T, M\}$. The baseline takes $\bar{V}_{st}$ and q_{st} as given to the receiver rather than deriving them from an information supplier's optimization problem. Nevertheless, they may differ across sources and change with the access regime. The model therefore accommodates supply-side changes in reduced form while keeping attention, processing, and knowledge formation at its centre.

2.4 How internet access changes knowledge

Let z_t denote the availability of internet access, corresponding to the change in digital infrastructure studied empirically. By contrast, a_t is the share of attention allocated to modern information sources. Greater access may increase a_t, but it may also expand the volume of online content or change its quality and perceived credibility. We therefore treat attention allocation as one of several channels through which internet access affects knowledge.

The marginal effects below are evaluated where $H_{t+1} > 0$ and the zero floor does not bind.[5] Because H_t is predetermined with respect to contemporaneous access, the marginal effect of each channel on H_{t+1} equals its marginal effect on ΔH_t. Consider first the attention margin:

$$\frac{\partial H_{t+1}}{\partial a_t} = A\left(\bar{V}_{Mt}Q_{Mt} - \bar{V}_{Tt}Q_{Tt}\right) - 2\lambda\Omega_t\left(\bar{V}_{Mt} - \bar{V}_{Tt}\right). \tag{7}$$

Rearranging equation (7) gives

[3] In practice, this ambiguity is likely limited because fixed-line home internet connections are rare in West Africa and mobile connectivity dominates (World Bank, 2024).

[4] The same convention applies to H_t, a_{st}, δ_{st}, Q_{st}, I_t, Ω_t, A, and λ, all of which retain their individual-specific interpretation. By contrast, information volume $\bar{V}_{st}$ and objective quality q_{st} are source-level objects common across individuals.

[5] If the expression inside the maximum in equation (4) is strictly negative, the next-period stock is locally constant at zero. At the boundary, the law of motion may not be differentiable.

$$\frac{\partial H_{t+1}}{\partial a_t} = \underbrace{\bar{V}_{Mt}\left(AQ_{Mt} - 2\lambda\Omega_t\right)}_{\substack{\text{net marginal return} \\ \text{to modern attention}}} - \underbrace{\bar{V}_{Tt}\left(AQ_{Tt} - 2\lambda\Omega_t\right)}_{\substack{\text{net marginal return} \\ \text{to traditional attention}}}. \tag{8}$$

The two terms in equation (8) are the net marginal returns to attention allocated to modern and traditional sources. A marginal shift toward modern media raises the next-period knowledge stock if and only if the net marginal return to modern attention exceeds the corresponding return to traditional attention. If the two sources supply the same volume per unit of attention, $\bar{V}_{Mt} = \bar{V}_{Tt}$, the shift raises next-period knowledge exactly when $Q_{Mt} > Q_{Tt}$. If modern media supply more content, the shift also raises the processing burden and is beneficial only when the informational advantage of modern media is sufficiently large. If modern media supply less content, the resulting reduction in processing costs may raise next-period knowledge even when their gross informational return is lower. Reversing these conditions gives the cases in which attention substitution lowers next-period knowledge.

The marginal effect of expanding the volume supplied by source s on the next-period knowledge stock is

$$\frac{\partial H_{t+1}}{\partial \bar{V}_{st}} = a_{st}\left(AQ_{st} - 2\lambda\Omega_t\right), \qquad s \in \{T, M\}. \tag{9}$$

Additional content raises next-period knowledge when its quality-adjusted learning value exceeds its marginal processing cost. For modern content, this condition is

$$AQ_{Mt} > 2\lambda\Omega_t. \tag{10}$$

The threshold is more likely to hold for individuals with greater processing ability, for higher-quality and more trusted content, and at lower initial levels of exposure. It generates an inverted-U relationship between information volume and the next-period knowledge stock without assuming that online content is intrinsically harmful.

Trust and objective quality produce the following margins:

$$\frac{\partial H_{t+1}}{\partial \delta_{st}} = A a_{st} \bar{V}_{st} q_{st}, \qquad \frac{\partial H_{t+1}}{\partial q_{st}} = A a_{st} \bar{V}_{st} \delta_{st}. \tag{11}$$

Greater trust raises next-period knowledge when a source is accurate and lowers it when the source is misleading. An improvement in objective quality raises next-period knowledge whenever the source receives positive attention and trust.

The total effect of internet access on the next-period knowledge stock combines these potentially conflicting margins. Holding A and H_t fixed, the chain rule gives

$$\frac{dH_{t+1}}{dz_t} = \frac{\partial H_{t+1}}{\partial a_t}\frac{da_t}{dz_t} + \sum_{s\in\{T,M\}}\left[\frac{\partial H_{t+1}}{\partial \bar{V}_{st}}\frac{d\bar{V}_{st}}{dz_t} + \frac{\partial H_{t+1}}{\partial \delta_{st}}\frac{d\delta_{st}}{dz_t} + \frac{\partial H_{t+1}}{\partial q_{st}}\frac{dq_{st}}{dz_t}\right]. \tag{12}$$

Equation (12) decomposes the effect of period-t access on H_{t+1}. Internet access raises the next-period knowledge stock when it shifts attention toward sources with higher effective returns, supplies useful content below the overload threshold, improves objective quality, or increases trust in accurate sources. It lowers the next-period stock when attention moves toward sources with lower effective returns, growth in volume pushes exposure beyond the threshold, source quality deteriorates, or trust moves away from accurate information. A negative value of dH_{t+1}/dz_t means that greater access lowers the period-$t+1$ stock relative to the counterfactual. Absolute knowledge declines between periods occur when $\Delta H_t < 0$, as characterized by equation (5). The model supports positive, zero, or negative effects. The AGIE examines information quality and source substitution under controlled conditions. The natural experiment then examines whether changes in established information sources, trust, and uncertainty accompany changes in health knowledge, following expanded internet access.

Objective quality may respond to the incentives facing information suppliers. Serra-Garcia (2026) finds that incentives to attract attention cause communicators to transmit less information. These incentives increase information demand among interested receivers but lower knowledge among the majority who do not request further detail. The result operates through omitted information rather than an increase in factually false statements. In theoretical work, Chen and Suen (2023) show that competition for attention can reduce news quality, while Acemoglu et al. (2024) show that engagement-maximizing platforms may create filter bubbles that weaken discipline on low-reliability content. These findings motivate allowing q_{Mt} to change across information environments even though the baseline model does not derive it from supplier behaviour.

2.5 Empirical implications and scope

The model organizes the empirical analysis around four predictions. Internet access should shift attention toward modern sources and may crowd out traditional ones. Where the zero floor does not bind, the marginal return to additional information falls as total exposure or the processing-cost parameter rises, holding all else fixed. Negative effects on next-period knowledge should be larger when modern information has lower objective quality. Finally, declining trust in accurate traditional sources provides a separate route through which access can reduce knowledge. Heterogeneity by education, literacy, age, or prior knowledge can reflect differences in A, although these characteristics may also affect attention and source evaluation.

The framework applies beyond health whenever individuals accumulate domain-specific knowledge from competing sources under an attention constraint. Aggregate knowledge can be obtained by summing individual knowledge stocks, although such an aggregation abstracts from social learning and other interactions across individuals.

3 The Agentic General Intelligence Experiment (AGIE)

The simplest explanation for a decrease in knowledge is unfortunately the most challenging to measure. Indeed, online claims may be lower in objective quality and individuals may shift attention towards them. However, until now, estimating this effect would require implausibly detailed data or an ethically ambiguous field experiment. Providing incorrect information to survey participants to measure decreasing health knowledge would rightfully prevent this experiment from running.

Our new Agentic General Intelligence Experiment provides a cost and time effective alternative for social scientists when a traditional randomised controlled trial (RCT) would be implausible. It assigns language-model agents to feeds of different quality while holding the model, respondent profiles, information volume, and outcome questions fixed. The experiment asks how information quality affects knowledge and whether replacing reliable traditional content with noisy online information can reduce maternal HIV knowledge. Its estimates are causal within the specified computational environment. The design attempts to mimic human behaviour through the characteristic profiles.

This design builds on work using language models as simulated economic agents. Horton et al. (2026) assign language-model agents endowments, information, and preferences, while Aher et al. (2023) construct synthetic samples to reproduce established laboratory findings. Instead, this application runs like an RCT. It draws respondent profiles from a weighted survey population, runs a fixed local model without live internet access, and assigns information treatments that vary the quality channel in the conceptual framework.

3.1 Model environment

The experiment runs locally through Ollama using the open-weight model Gemma 4 E4B.[6] The primary specification sets temperature to 0.3, top-p to 0.9, and the context window to 8,192 tokens. Running a fixed local model has two major benefits. First, it prevents live web retrieval by the model. This would be particularly problematic if the experiment followed a similar structure to a published experiment, as the model could draw from their results to determine responses. Second, it allows the model version to be archived with the replication materials. This is not possible with an online model like ChatGPT that does not archive older versions of their model.

[6]The archived checkpoint is listed as `gemma4:e4b-it-q4_K_M`.

A language model generates an answer from a probability distribution over tokens. Temperature controls how concentrated that distribution is, while top-p restricts sampling to the candidate tokens carrying most of its probability mass. The primary settings permit some response variation but favour high-probability answers. This is deemed appropriate for a health related questionnaire. Figure 1 illustrates these parameters. In Appendix A, we regenerate the experiment at temperature 0.5 and 0.7, holding all else fixed. Reassuringly, the results remain very similar.

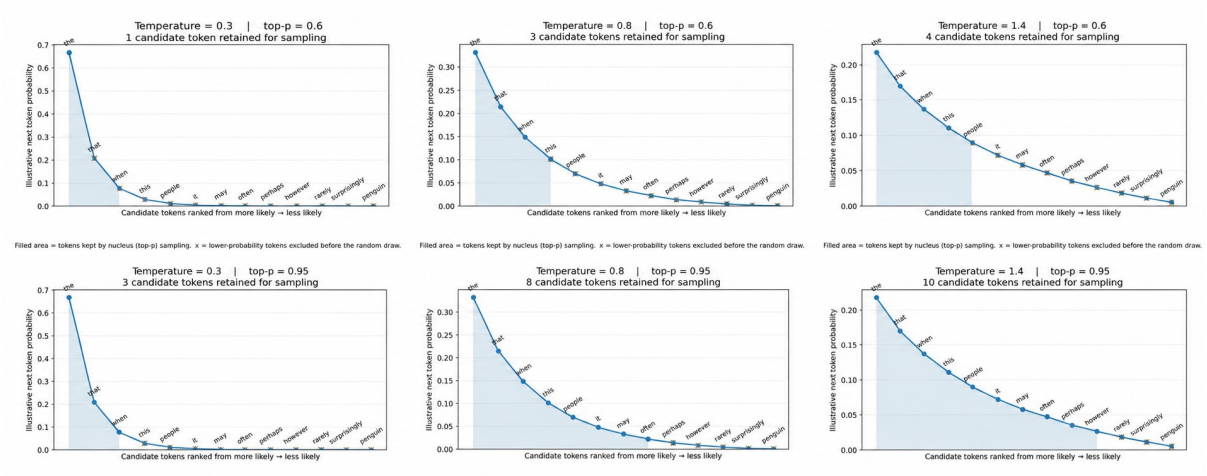


Figure 1: Intuition for temperature and top-p

The system prompt instructs the model to answer as the described respondent, using the profile, any information supplied in that call, and ordinary prior knowledge consistent with the profile. It omits treatment status, scenario names and any other design artefacts to ensure no internal bias occurs. Baseline and endline answers are elicited in separate calls, and agents never see previous scenario's conversation history or other agent's responses. This ensures conversational memory does not bias the results.

3.2 Experimental design

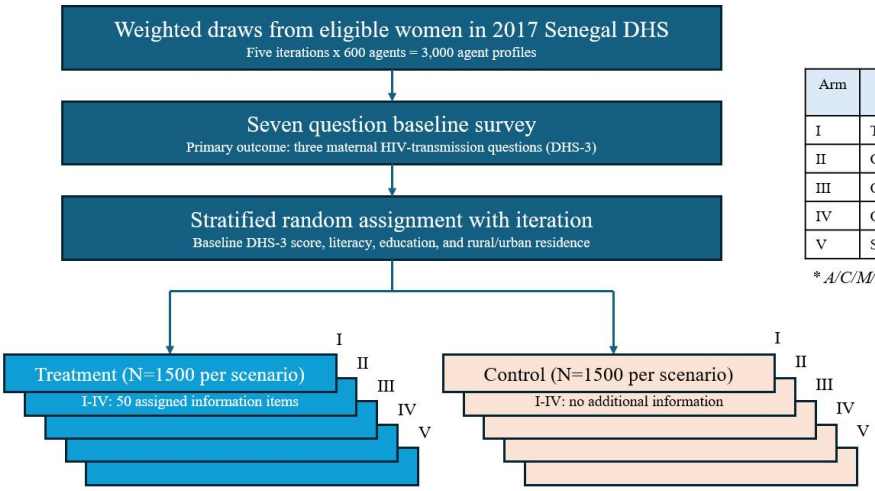
The sampling frame consists of women in the 2017 Senegal DHS who report no baseline internet use. Each profile records age, region, rural or urban residence, education, literacy, wealth, marital status, media access, antenatal-care exposure, health-worker contact, and maternal HIV transmission answers. Five iterations of 600 profiles produce 3,000 synthetic agents.

Before assignment, every agent answers seven HIV knowledge questions. The sampling frame and outcome questions are designed to mimic the DHS survey as it is used in a complementary natural experiment to support the agentic results. The agents use the sampled respondents' real answers to the 3 DHS questions as their baseline scores. Then, they answer the four additional questions with their unique profiles. Using each agent's observed maternal HIV answers is a strong benefit to the sample design as it gives agents

more context to answer the additional four questions; in practice this reduces the AI model's reliance on training data which prevents unrealistic baseline knowledge.

The primary outcome is the fraction answered correctly across the three DHS questions on transmission during pregnancy, delivery, and breastfeeding. A secondary index adds four questions: whether a healthy-looking person can have HIV, whether mosquitoes transmit HIV, whether HIV has a cure, and whether it can be transmitted by kissing. The correct answers to these four questions are yes, no, no, and no.[7] Both indices range from zero to one, and we also estimate effects for each question separately.

The model tests two separate response styles to determine best practice for future AGIEs. Half of the sample is told to directly respond with yes, no, or don't know, and the other half is encouraged to give a conversational response. The conversation (reasoning) response aims to emulate the human behaviour of internal reasoning/monologue. However, it does come at a serious computational cost. The 'don't know' responses are coded as incorrect answers in this experiment.



Arm	Information Environment	A/C/M/I (%)*
I	Traditional Media, High Quality	70/10/0/20
II	Online, Informative	50/15/15/20
III	Online, Mix	30/30/20/20
IV	Online, Noisy	20/30/30/20
V	Switching from I to IV	

** A/C/M/I – accurate/confusing/misleading/irrelevant*

Figure 2: Experimental sequence

Figure 2 describes the interventions. In scenarios I-IV, treated agents receive 50 items and control agents receive no new information before the endline survey. Each treated agent receives a separately sampled feed with the category shares specified for that scenario.

Scenario I provides a high-quality traditional-media benchmark, emulating an information environment that could exist without internet access. Scenarios II–IV use online-style items and progressively replace accurate material with confusing or misleading content. We treat Scenario IV as the preferred online environment, reflecting the previously stated supply-side incentives for sensationalised content alongside the growth of 'vagueposting', polarisation, and fake news on social media. This choice is consistent with evidence from Suarez-Lledo and Alvarez-Galvez (2021), whose systematic review of 69 studies reports

[7]The cure question is scored under the intended interpretation of a cure available through routine HIV care. Appendix A.3 discusses the scope for an 'exceptional-case' interpretation.

misinformation prevalence rates of 30% for medical treatments, 40% for diseases, and 43% for vaccines.

The message library applies fixed categories to information. Accurate items state verified guidance; confusing items are incomplete or ambiguous; misleading items contain a false claim or consequential omission; and irrelevant items contain no HIV information. There are equal numbers of messages containing information for each of the seven health questions. The information library uses a variety of different language and sentence structures to imitate different sources. For example, plain and straightforward language for an official health report, or conversational styles for social media comments. This variation aims to emulate a real heterogeneous information environment. The appendix A contains the full information feed library, and the experiment's python script is available upon request.

3.3 Randomisation and estimation

Assignment occurs after the baseline survey. Within each iteration, agents are stratified by their baseline three-question score, literacy, education, and rural or urban residence, then allocated in approximately equal numbers to treatment and control within each stratum. Every iteration contains 300 treated and 300 control agents. The experiment reuses the same sampled profiles and treatment assignments across Scenarios I-V. However, each scenario is evaluated independently in a fresh, stateless model call containing only that scenario's information environment. This means that earlier exposure and responses are not supplied to the model, so the later scenarios do not turn cumulative. Estimates across scenarios are correlated only through the use of the same profile generated agents.

Table 1 confirms that assignment produced comparable groups. Control and treatment agents both score 0.635 on the baseline three-question index, and 0.612 and 0.606 on the seven-question index. Differences in age, residence, education, literacy, and media access are small and none are statistically significant at conventional levels.

Table 1: Baseline Balance

	Control mean	Treatment mean	Difference	t-stat
Age	29.711	29.474	-0.237	(-0.67)
Rural	0.605	0.612	0.007	(0.41)
Any education	0.373	0.373	0.000	(0.00)
Can read at least partly	0.315	0.317	0.001	(0.08)
Radio access	0.711	0.711	0.000	(0.00)
TV access	0.538	0.513	-0.025	(-1.39)
Direct response protocol	0.500	0.500	0.000	(0.00)
Baseline DHS-3 score	0.635	0.635	-0.000	(-0.01)
Baseline full-seven score	0.612	0.606	-0.007	(-0.68)
Observations	1,500	1,500		

Notes: The three DHS-3 answers are observed in the sampling frame and the four additional answers are elicited before treatment assignment. Treatment is randomised within DHS-3 score, literacy, education, and rural or urban residence. Difference is treatment minus control; t statistics are in parentheses.

For each scenario s and outcome k, we estimate the ANCOVA regression:

$$Y_{isk}^{post} = \alpha_{sk} + \tau_{sk} T_i + \rho_{sk} Y_{ik}^{base} + \lambda_{sk,r(i)} + X_i' \gamma_{sk} + \varepsilon_{isk}, \tag{13}$$

where i indexes agents, s scenarios, and k outcomes. T_i indicates assignment to the feed, Y_{ik}^{base} is the corresponding baseline outcome, and $\lambda_{sk,r(i)}$ denotes iteration fixed effects. The vector X_i contains additive controls for literacy, education, and rural or urban residence. The coefficient τ_{sk} is the intention-to-treat effect on endline knowledge. Heteroskedasticity-robust standard errors reflect variation across agents, including their independently sampled feeds and model responses, conditional on the fixed message library, respondent profiles, and model settings.

3.4 Experimental results

Average treatment effects. Table 2 reports the ANCOVA estimates separately by response protocol and for the pooled sample. We use the pooled estimates as the main summary. Relative to receiving no new information, the high-quality traditional-media feed raises the three-question DHS score by 0.194 (SE 0.013). The corresponding effects are 0.165 (SE 0.012) for the informative online feed, 0.118 (SE 0.012) for the mixed online feed, and 0.078 (SE 0.012) for the noisy online feed. All four estimates are statistically significant at the 1 per cent level. The point estimates therefore decline monotonically as accurate material is replaced by confusing and misleading content, although even the noisiest feed improves maternal-transmission knowledge relative to providing no information.

Table 2: ANCOVA Treatment Effects by Response Protocol

	Direct Response		Reasoning Response		Pooled Response	
	DHS-3	Full-7	DHS-3	Full-7	DHS-3	Full-7
	(1)	(2)	(3)	(4)	(5)	(6)
Scenario I: Traditional media	0.187***	0.211***	0.202***	0.182***	0.194***	0.197***
	(0.018)	(0.011)	(0.018)	(0.010)	(0.013)	(0.007)
Scenario II: Online, informative	0.167***	0.166***	0.162***	0.123***	0.165***	0.144***
	(0.018)	(0.011)	(0.017)	(0.010)	(0.012)	(0.007)
Scenario III: Online, mixed	0.113***	0.082***	0.124***	0.047***	0.118***	0.065***
	(0.018)	(0.011)	(0.017)	(0.009)	(0.012)	(0.007)
Scenario IV: Online, noisy	0.081***	0.001	0.075***	-0.047***	0.078***	-0.023***
	(0.017)	(0.010)	(0.016)	(0.009)	(0.012)	(0.007)
Scenario V: Media substitution, I to IV	-0.094***	-0.219***	-0.117***	-0.236***	-0.107***	-0.228***
	(0.017)	(0.010)	(0.017)	(0.010)	(0.012)	(0.007)
Observations	1,500	1,500	1,500	1,500	3,000	3,000

Notes: Cells report HC2-robust ANCOVA estimates; standard errors are in parentheses. Columns (1)–(2) use the direct-response protocol, columns (3)–(4) use the brief-reasoning protocol, and columns (5)–(6) pool both protocols. Scenarios I–IV estimate assignment to the stated information environment relative to no new information, conditional on baseline knowledge, iteration, literacy, education, and rural or urban residence. Scenario V estimates the traditional-to-online switch relative to continued traditional media, conditional on the post-traditional outcome and the same controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The full-seven index is more sensitive to deteriorating content. The pooled effect falls from 0.197 under traditional media to 0.144 under the informative online feed and 0.065 under the mixed feed. Under the noisy online environment, it becomes negative at −0.023 (SE 0.007). The same ordering appears under both response protocols. One difference arises in Scenario IV: its effect on the full-seven index is close to zero under direct responses, at 0.001 (SE 0.010), but negative under reasoning, at −0.047 (SE 0.009). The results for the three-question DHS outcome are less sensitive to response protocol.

Scenario V changes the relevant counterfactual from no new information to continued access to traditional media. All agents first receive the traditional-media feed and complete an intermediate knowledge assessment. Control agents then receive another traditional-media feed, while treated agents switch to the noisy online environment. Conditional on knowledge after the common first stage, the switch reduces the pooled three-question score by 0.107 (SE 0.012) and the full-seven score by 0.228 (SE 0.007). Both estimates are statistically significant at the 1 per cent level and are similar under the direct and reasoning protocols.

Table 3 reports the corresponding pooled means. Among agents who continue receiving traditional content, the three-question score rises from 0.808 after the first stage to 0.866 after the second. Among agents who switch to the noisy online feed, it falls from 0.825 to 0.759. The full-seven score follows the same pattern: it rises from 0.797 to 0.857 under continued traditional media but falls from 0.801 to 0.629 following the switch. The negative Scenario V effect therefore reflects an absolute deterioration among switchers, rather than a smaller improvement relative to continued traditional media. The continued improvement following a second round of traditional content suggests repeated delivery can reinforce knowledge. This pattern reinforces experimental evidence on repeated health communication, particularly when messages are accompanied by opportunities to retrieve or apply the information (Cepeda et al., 2008; Betts et al., 2019; Mo et al., 2014).

Table 3: Knowledge Before and After Media Substitution in Scenario V

Stage	Group	*N*	DHS-3 mean	Full-seven mean
Post Traditional	Control	1,500	0.808	0.797
Post Traditional	Treated	1,500	0.825	0.801
Post Switch	Control	1,500	0.866	0.857
Post Switch	Treated	1,500	0.759	0.629

Notes: Means are average shares correct across the DHS-3 and full-seven outcome sets. The table pools the direct-response and brief-reasoning protocols. "Post Traditional" records knowledge after all agents receive the traditional-media feed. "Post Switch" records knowledge after treated agents switch to the noisy online feed while control agents continue receiving traditional-media content.

Scenarios IV and V clarify why the choice of counterfactual matters. The noisy online feed can improve maternal-transmission knowledge when the alternative is no additional

information. It nevertheless produces a substantial loss when it replaces repeated exposure to reliable traditional content.

Question-level effects. Appendix Table 22 reports effects for the seven individual questions. The largest losses occur for the two questions that directly confront common transmission myths. Under the noisy online feed, knowledge that mosquito bites do not transmit HIV falls by 0.259, while knowledge that kissing does not transmit HIV falls by 0.324. When agents switch from traditional media to the noisy online feed in Scenario V, the corresponding effects are −0.392 and −0.589. All four estimates are statistically significant at the 1 per cent level.

Several of the more direct factual questions remain positively affected in Scenario IV. Knowledge of transmission during pregnancy and delivery rises by 0.104 and 0.152, respectively, relative to no new information, while the estimate for breastfeeding transmission is −0.022 and statistically insignificant. Under Scenario V, the effects are negative for six of the seven questions. Replacing reliable traditional content with the noisy feed therefore produces a broader deterioration than exposure to that feed against a no-information counterfactual.

Stability and sensitivity. Appendix Table 21 reports estimates separately for the five iterations. The sign of every scenario effect is unchanged across iterations. For the three-question outcome, the Scenario IV coefficients range from 0.060 to 0.097, while the Scenario V coefficients range from −0.122 to −0.073. For the full-seven outcome, the corresponding ranges are −0.037 to −0.008 and −0.243 to −0.211. The principal findings are therefore not driven by one sample of profiles or one set of model calls.

A robustness exercise is done with the temperature of the model, which compares the reasoning response protocol at temperatures 0.3, 0.5 and 0.7. This keeps the same agent profiles from the main specification to isolate the effects of changing model temperature only. Appendix Table 23 shows that the ordering of the scenarios remains unchanged. For example, the Scenario V effect is −0.117 at temperature 0.3, −0.120 at 0.5, and −0.129 at 0.7, for the three-question outcome. The corresponding full-seven estimates are −0.236, −0.244 and −0.233 , respectively. The results are therefore insensitive to the choice between these three sampling temperatures.

3.5 Interpretation, scope and discussion

The experiment isolates the information-quality channel represented by q_{Mt} in the conceptual framework. Within the environments considered, the traditional-media feed produces the largest positive point estimate for maternal HIV knowledge. An informative online environment also improves knowledge, but the gain becomes progressively smaller as confusing and misleading content occupies more of the feed.

The appropriate conclusion depends on the counterfactual. Scenario IV shows that even a noisy online feed can convey useful maternal-transmission facts when compared with a no information state. Scenario V tests the source-substitution channel (a_t) in the conceptual framework: agents move from repeated exposure to reliable traditional content to a noisier online environment. That switch reduces the three-question score by 0.107. Because Scenario V imposes complete substitution, it provides a limiting-case benchmark for the consequences of replacing one source with another. A setting in which people retain access to established sources after gaining internet access would involve a less complete change in their information environment.

The experiment also permits a treatment that would be difficult to assign to human participants. Deliberately exposing people to false health claims would raise serious ethical concerns. Synthetic agents make it possible to examine how such content affects measured knowledge without placing participants at risk. AGIEs may therefore be useful for piloting costly field experiments, testing questionnaires and analysis plans, and studying treatments that cannot ethically be assigned to people.

Memory and attention remain important limitations. The model receives the entire feed within its context window and may compare its contents more systematically than a person would. Human respondents can forget most of what they encounter while retaining claims that are salient, controversial, or compatible with their prior beliefs. The current design neither imposes selective recall nor varies the number of items. It therefore cannot identify the information-overload term, $\lambda_i \Omega_t^2$, in the conceptual framework. Further research could aim to incorporate memory and attention through explicit attention budgets, delayed recall, or probabilistic forgetting.

The model also brings prior knowledge about HIV from its training data. In this experiment we choose to anchor the three DHS questions to respondents' observed answers. Anchored agents answer the four additional questions correctly less often than unanchored agents.[8] Of course, a short DHS profile cannot reproduce a respondent's lived experience.

Finally the results show promising consistency across iterations, temperature and response protocols. Future validation should compare synthetic and human responses under the same experimental conditions to confirm AGIE accuracy across different specification. Until that evidence is available, the experiment provides controlled evidence about how these language-model agents respond to assigned information, not necessarily how human participants would respond to an identical feed.

[8] In exploratory runs without anchoring, correct-response rates on the seven questions ranged from approximately 85 to 95 per cent. With anchoring, the rate on the four additional questions was approximately two-thirds, close to the average for the three maternal questions in the DHS.

3.6 Reporting and reproducibility

All preferred specifications use the full randomised sample of 3,000 agents. The replication package contains the inputs, raw outputs, and code needed to reconstruct treatment assignment, response coding, and estimation, together with the model checkpoint and prompt settings. Treatment labels, hypotheses, and expected signs are excluded from all agent-facing prompts (Sharma et al., 2024).[9]

4 Natural Experiment

The Agentic General Intelligence Experiment shows how information quality and environment substitution can change measured knowledge within a language model. We now examine whether expansions to internet infrastructure, which plausibly change an individual's information environment, impact women's maternal HIV knowledge. The staggered rollout of fibre infrastructure allows us to estimate a reduced-form access effect. It also validates whether source substitution actually occurs following internet access.

4.1 Setting

Fibre-optic cables form the backbone of high-speed internet networks. They carry large volumes of data across regions and connect to local mobile towers and distribution networks. Households can benefit from this backbone without having fibre lines directly in their homes. This network structure is particularly relevant in West Africa, where mobile data provide the main form of internet access (World Bank, 2024). Mobile connections rely on fibre backhaul to connect towers to the wider internet.

Submarine cable landings between roughly 2009 and 2012 expanded the capacity available to the region. Operators subsequently extended terrestrial routes inland, connecting communities and increasing the capacity available to local networks. Earlier systems supported 2G connectivity, calls and text messaging, but offered virtually no scope for regular online search and social media due to extremely slow transmission speeds (ITU, 2001). The empirical variation comes from differences in when inland terrestrial cables reached different regions.

Internet use grew substantially over the sample period. Between 2012 and 2024, the share of the population using the internet rose from 9.2 to 33.6 percent in Sub-Saharan Africa and from 73.7 to 93.0 percent in high-income economies—relative increases of 265 and 26 percent, respectively (World Bank, 2026). The expansion examined here therefore occurred during the transition from limited access to regular digital use. In a region where

[9]Further implementation details are documented in the replication README.

mobile connections dominate, terrestrial fibre measures the backbone serving the towers and local networks through which households are most likely to connect.

Engineering and commercial considerations shaped the rollout. Routes commonly followed roads, and operators had stronger incentives to connect cities and areas with greater expected demand. Network nodes (or exchange points) determined where the backbone could branch into local infrastructure and broadband towers. A cable passing through a district may provide little local service if there is no suitable junction. The treatment definition therefore incorporates proximity to both cables and nodes, unlike cable-only measures used in related work, and the empirical checks vary the spatial rules used to assign exposure.

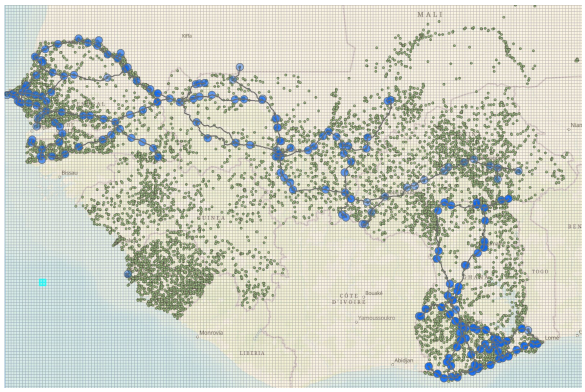
Prior to widespread internet access, information on maternal health, HIV prevention, and family planning was disseminated through established channels such as health facilities, community outreach programmes, and broadcast media (Iacoella et al., 2022b). Radio in particular played a prominent role in reproductive-health communication across Sub-Saharan Africa, and exposure to traditional media has been shown to be strongly associated with health knowledge; in a study covering 31 African countries, radio exposure accounted for 26.1% of the explained variation in family-planning knowledge (Mutumba, 2022). Internet expansion changed the information environment in both the quantity of available content and the range of different opinions.

4.2 Data and measurement

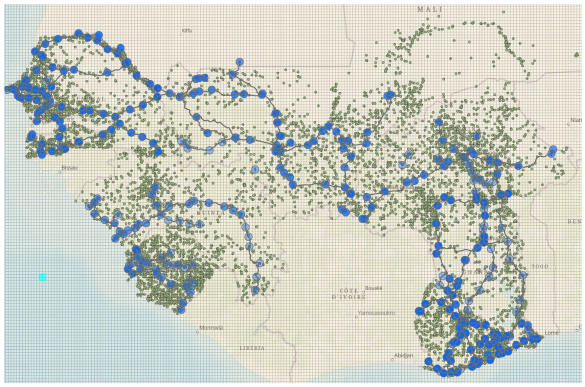
The analysis combines DHS repeated cross sections, annual maps of terrestrial fibre infrastructure, and Afrobarometer repeated cross sections. The DHS supplies the knowledge outcomes and reproductive-health measures. The network maps determine the location and timing of treatment. Afrobarometer measures internet frequency, news sources, and trust in political and media institutions.

The main sample covers Burkina Faso, Ghana, Guinea, Mali, Senegal, and Sierra Leone from 2003 to 2023, with exact coverage varying by country and outcome. A country enters the sample when the fibre maps record meaningful within-country rollout variation and the DHS contains at least two usable rounds with the relevant maternal HIV questions. [10] The unit of observation is an individual respondent. The women's dataset contains 227,544 observations. After accounting for outcome availability, complete controls, and the road restriction, the preferred pregnancy and maternal-index specifications contain 163,785 and 167,868 women, respectively.

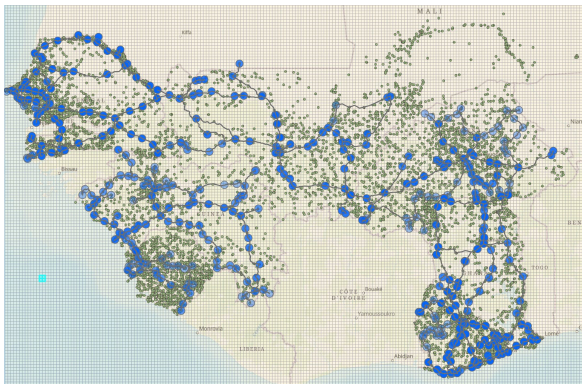
[10]For example, Liberia has several viable DHS rounds but little recorded fibre expansion during the relevant period, while Togo has substantial fibre-optic expansion but only one DHS round in the sample window. Nigeria has a vast fibre network and multiple DHS rounds, but it was not possible to confirm spatial connectivity at a fine enough level to validate treatment. Sierra Leone is not present in Afrobarometer analysis due to data request limitations.



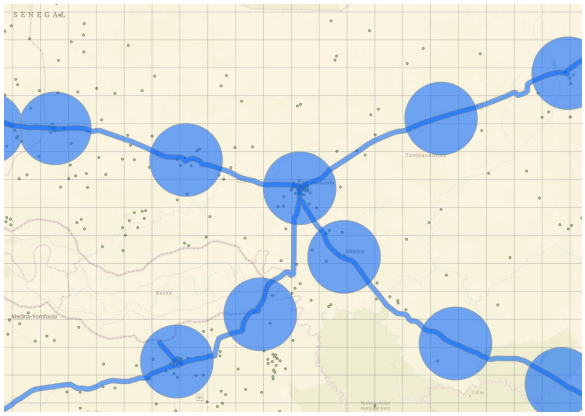
(a) 2013



(b) 2018



(c) 2023



(d) Zoomed view

Figure 3: Expansion of the fibre-optic network. Panels a–c show the network in 2013, 2018, and 2023. Panel d provides a detailed view of the highlighted area.

Women form the benchmark sample. The subject concerns mother-to-child transmission, and women appear more consistently in the relevant DHS modules. Full-sample and male estimates are reported in Appendix B. The male sample provides a useful comparison but is smaller and drawn from fewer survey rounds.

Respondents are asked whether HIV can pass from mother to child during pregnancy, during delivery, and through breastfeeding. The three route-specific outcomes are coded zero for an incorrect answer, one for "don't know", and two for the verified answer. [11] The pregnancy question is the benchmark outcome because it concerns transmission over the broadest period before birth. The maternal HIV index used in the paper is the pre-existing four-item `aids_maternal` measure. It averages the general mother-to-child-transmission item with binary indicators for correct answers to the pregnancy, delivery, and breastfeeding questions. Separate regressions for "don't know" answers distinguish uncertainty from stated incorrect beliefs.

An incorrect answer records a stated belief that conflicts with verified guidance, whereas a "don't know" response records uncertainty over which answer is correct. Both lower the benchmark knowledge measure but imply different things about the information environment available – uncertainty or misinformation. The decomposition of response regressions shows whether treatment shifts respondents toward a false answer, uncertainty, or a combination of the two.

These questions measure a specific form of health knowledge with direct implications for preventive behaviour. Knowledge that HIV can be transmitted from mother to child during pregnancy is a key component of the prevention of mother-to-child transmission (PMTCT) cascade, as it can affect whether women seek HIV testing during pregnancy, initiate antiretroviral therapy following a positive diagnosis, and adhere to medical guidance designed to reduce vertical transmission (World Health Organization, 2024b,a). These behaviours are consequential: in the absence of intervention, mother-to-child transmission rates range from 15% to 45%, whereas effective PMTCT interventions can reduce transmission to below 5%. Knowledge of transmission during delivery and breastfeeding captures awareness of additional transmission risks that remain relevant around and after birth.

The DHS also records whether respondents encountered family-planning messages through radio, television, and newspapers or magazines. Additional outcomes include contact with family-planning or health workers and the number of antenatal visits. These measures capture channels through which reproductive-health advice reached households before and during the expansion of online access.

Annual Africa Bandwidth Maps (2026) provide the location and timing of terrestrial fibre routes and network nodes. We digitise the network in ArcGIS and cross-validate routes against AfTerFibre, Infrapedia and the International Telecommunication Union's

[11] The verified answer to all three route-specific questions is "yes".

Broadband Maps to confirm where coverage overlaps (AfTerFibre, 2026; Infrapedia, 2026; ITU, 2026).

A respondent is considered treated when her enumeration area lies within 1 kilometre of a fibre cable or 15 kilometres of a network node in the survey year. The narrow cable buffer allows for a short branch from the main route to cover a village with a mobile tower. The wider node buffer reflects the distribution role of those junctions, where the backbone can split into smaller branches to connect broadband towers and in-line connections. Nodes also house equipment that maintains signal strength over long distances. [12] The estimated scope of this connectivity is approximately 15 kilometres. Alternative robustness specifications use a 15-kilometre node-only treatment, a 10-kilometre node radius and 1-kilometre cable radius, and inverse distance from routes and nodes. The preferred binary treatment switches on in the recorded connection year: a respondent is treated when the survey year is greater than or equal to that year.

All main DHS specifications use the road-restricted sample defined in Section 4.3.

Treatment records proximity to internet-enabling infrastructure. Some households near fibre remain offline, whereas people farther away may connect at work or when travelling to nearby networks. Therefore, the DHS coefficients give the reduced-form effect of infrastructure exposure. Afrobarometer provides a separate estimate of the relationship between the same treatment and reported use.

DHS displaces geographic coordinates to protect respondent confidentiality. The displacement is up to 2 kilometres in urban areas and up to 5 kilometres in rural areas, but enumeration areas remain in their respective district when they are near a border. Enumeration areas close to a treatment boundary may therefore be misclassified. Sensitivity to this potential misclassification is assessed using alternative treatment measures. In particular an inverse-distance measure, equal to one for enumeration areas next to the cable network and approaching zero as distance increases, is used. This produces estimates of the same negative sign as the baseline result.

Afrobarometer has a smaller sample but covers more survey rounds than the DHS; Appendix B.1 reports coverage by country and wave. Its control vector covers the same broad domains as the DHS specification using the closest available variables. For example, the frequency of going without food or cash income proxies for household economic circumstances because Afrobarometer lacks the DHS wealth index. These outcomes provide proxies for changes in attention to different media and trust in public information sources over the sample period.

[12] Signal loss over distance is known as attenuation. Node placement can therefore reflect engineering requirements as well as local demand.

4.3 Empirical design

We compare changes in maternal HIV knowledge within small geographic areas as the fibre network expands. Respondents are assigned to 10-kilometre by 10-kilometre grid cells, and the baseline specification is

$$Y_{iegt} = \beta FC_{et} + X'_{iegt}\lambda + \phi_g + \tau_{r(g)t} + \varepsilon_{iegt}, \tag{14}$$

where i indexes respondents, g grid cells, $r(g)$ the region containing grid cell g, e the surveyed enumeration area, and t survey years.[13] The indicator FC_{et} equals one once a surveyed location meets the preferred fibre-proximity rule. The vector X_{iegt} contains respondent, household, and geographic controls. Grid-cell fixed effects ϕ_g absorb time-invariant local characteristics, and region-by-year fixed effects $\tau_{r(g)t}$ absorb shocks common to respondents in the same ADM1 region and survey year, for example a region specific health campaign. Standard errors are clustered by grid cell. The coefficient β is the reduced-form effect of local infrastructure exposure on maternal HIV knowledge.

We introduce the controls sequentially to show how the treatment estimate changes as covariates are added. Demographic variables include age, marital status, and household size. Socioeconomic controls cover literacy, education, radio and television ownership, and wealth. The full specification adds urban residence and electricity; full-sample specifications also include gender. The sequence shows how the coefficient changes as observed respondent and household composition is held fixed.

Because the data are repeated cross sections, these controls serve a second practical role. A grid cell may contain respondents with different education, wealth, or household assets in successive survey rounds. The sequential specifications reveal whether such changes accompany the estimated treatment effect.

Identification relies on parallel changes in treated and comparison grid cells in the absence of fibre expansion. The road restriction retains every observation inside a fibre-proximity treatment zone, including its pre-treatment observations, and limits potential controls outside those zones to enumeration areas within 10 kilometres of a major road. Because fibre routes frequently follow transport corridors, this restriction makes treated and comparison locations more similar in road access and avoids using extremely remote rural areas as counterfactuals. Locations eventually reached by fibre are initially more urban, wealthier, and better educated, but grid-cell fixed effects absorb these persistent level differences. The remaining concern is differential local trends: fibre routes may reach growing cities, transport corridors, or areas receiving concurrent investments that also affect maternal HIV knowledge.

[13] The grid-cell specification is consistent with Hjort and Poulsen (2019), who use a variation of this cable-treatment design.

The benchmark addresses this concern in two ways. The road restriction narrows the pool of potential controls from the outset, and region-by-year fixed effects absorb regional changes in health campaigns and economic conditions. Unrestricted estimates show how much the sample restriction matters; city exclusions, maximum-distance restrictions, and alternative time fixed effects address related forms of differential growth. A stacked event study examines pre-treatment movement, and placebo outcomes assess whether rollout coincided with broader local changes. The estimates remain negative across these exercises, and none of the seven placebo coefficients is statistically significant.

The staggered rollout also complicates the conventional two-way fixed-effects (TWFE) estimate. Previously treated cells can enter the comparison group for cells connected later, and heterogeneous effects across cohorts or event time can produce difficult-to-interpret weights (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). The preferred specification for addressing staggered treatment timing is therefore a weighted clean-control stacked regression estimated on the road-restricted sample. For each treatment cohort, the stack retains an eight-year window on either side of connection and compares that cohort only with never-treated cells or cells treated after the end of the stack. Already-treated cells never serve as controls.

Treated observations receive unit weight, while control observations in stack g receive weight $w_g = N_g^T/N_g^C$, where N_g^T and N_g^C denote the numbers of treated and control observations in that stack. This weighting gives the treated and control groups equal total weight within each stack. The regression absorbs stack-by-grid and stack-by-region-by-year fixed effects. With heterogeneous effects, the pooled coefficient is a regression-weighted average of the cohort- and event-time-specific effects represented in the stacked DHS sample, rather than an equally weighted average across cohorts or event times. Appendix B.11 shows that replacing the baseline weights with the repeated-cross-section (Q)-weights proposed by Wing et al. (2024) leaves the estimates unchanged at the reported precision.

Given the broadly similar headline knowledge and subgroup results, we retain TWFE as a common reporting benchmark. Appendix Table 41 compares the mechanism and heterogeneity estimates with clean-control stacked DiD.

The Afrobarometer treatment-validity regression moves both sets of fixed effects up one geographic level:

$$InternetUse_{irct} = \alpha FC_{irct} + X'_{irct}\lambda + \mu_{cr} + \eta_{ct} + u_{irct}. \tag{15}$$

Here c indexes countries, r country-qualified regions, and t survey years. Country-region fixed effects μ_{cr} absorb time-invariant regional conditions, while country-by-survey-year fixed effects η_{ct} absorb national shocks in each survey year. Standard errors are clustered

by country-qualified town or village.[14] Reported use ranges from zero for no internet use to four for daily use. Equation 15 tests whether mapped fibre exposure predicts reported internet use. A positive estimate of α supports interpreting the treatment as an internet-access shock.

4.4 Baseline comparability and treatment validity

Table 4 compares eventual treatment groups before rollout. In the road-restricted sample, treatment-zone observations have higher raw pregnancy-transmission knowledge and characteristics associated with access to formal health information, including greater literacy, education, radio ownership, wealth, and urban residence. These level differences make the within-grid comparison central to the design. In a simple comparison of levels, this positive selection would favour treated areas. The negative within-grid estimate therefore runs counter to the baseline imbalance. If the same selection also predicts faster untreated growth in knowledge, any residual bias would be upward and would attenuate the estimated decline.

These starting points leave differential trajectories as the relevant selection concern. We therefore read the balance table together with the pre-treatment event-study coefficients: the table describes initial levels, and the event study examines movement before the mapped connection date. The stacked event study indicates that the pregnancy coefficients before connection do not show a sustained negative pattern.

Table 4: Baseline Balance

	Control mean	Treated mean	Raw difference	Adjusted difference	Standard error	*p*-value	*N*
Six-country road sample							
Pregnancy-transmission knowledge	1.451	1.499	0.048	0.033	0.019	0.078	48754
Age	29.080	27.918	-1.162	-0.997	0.139	0.000	48755
Married	1.586	1.284	-0.302	-0.225	0.021	0.000	48755
Household size	8.470	9.151	0.681	0.055	0.235	0.814	48755
Literacy	0.325	0.741	0.416	0.304	0.031	0.000	48600
Education	1.632	3.771	2.139	1.470	0.142	0.000	48541
Radio	0.711	0.784	0.072	0.074	0.014	0.000	48751
Wealth	2.595	3.703	1.108	0.905	0.075	0.000	48755
Rural	1.836	1.354	-0.482	-0.479	0.032	0.000	48755

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. Treated observations are enumeration areas that enter the preferred fibre-proximity treatment. Adjusted differences use the balance-regression specification recorded in the analysis output. Decimal-valued entries are rounded to three decimal places.

Table 5 confirms fibre proximity predicts reported internet use. The average effect is an increase of 0.191 on the five-point frequency scale. Columns 2 and 3 confirm this

[14] Afrobarometer has far fewer observations within repeatedly sampled local areas than the DHS, so the grid and region-by-year fixed effects are scaled up to region and country-by-year fixed effects, respectively. Using region-by-year fixed effects instead, without a time invariant grid cell fixed effect, produces a nearly identical treatment estimate.

relationship is strongest with among who have cash incomes. This is unsurprising as those groups are most able to purchase internet enabling technology. This relationship also holds with a female-only subsample (Appendix Table 27). The first-stage relationship supports interpreting mapped fibre proximity as a shock to local internet access.

Table 5: Fibre Proximity and Internet Use

	Average effect	Without Cash Income	Implied effect Subgroup total
Fibre proximity	0.191***		
	(0.0209)		
Fibre proximity: never (reference)		0.288***	0.288***
		(0.0374)	(0.0374)
Fibre proximity × once or twice		0.108**	0.395***
		(0.0548)	(0.0495)
Fibre proximity × several times		0.00618	0.294***
		(0.0424)	(0.0328)
Fibre proximity × many times		-0.199***	0.0886***
		(0.0421)	(0.0274)
Fibre proximity × always		-0.283***	0.00484
		(0.0446)	(0.0307)
Controls			
Age	-0.0144***	-0.0144***	
	(0.000476)	(0.000475)	
Female	-0.307***	-0.307***	
	(0.0124)	(0.0123)	
Education	0.253***	0.253***	
	(0.00396)	(0.00396)	
Without food	-0.0317***	-0.0333***	
	(0.00555)	(0.00558)	
Without cash income	-0.0516***		
	(0.00584)		
Rural	-0.339***	-0.337***	
	(0.0217)	(0.0216)	
Country-region FE	Yes	Yes	
Country-by-year FE	Yes	Yes	
Mean dependent variable	0.853	0.853	
Observations	38,831	38,831	
R^2	0.437	0.439	

Notes: Internet use, measured from zero (never) to four (every day). "Fibre proximity" is the treatment indicator derived from the mapped fibre network. Levels 0–4 refer to the frequency of going without cash income in the past 12 months. Column (1) reports the average-effect specification, with a linear without cash income control. Column (2) includes without cash income category indicators and their interactions with fibre proximity. Both specifications include region and country-by-year fixed effects. Standard errors, clustered by town/village, are in parentheses; subgroup totals use delta-method standard errors. Stars use the reported regression t-test p-values in columns (1)–(2) and the reported z-test p-values for subgroup totals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.5 Maternal HIV knowledge

Table 6 reports the main DHS estimates. In the preferred road-restricted specification for women, fibre proximity reduces knowledge of transmission during pregnancy by 0.099 points on the 0–2 scale. Knowledge of transmission during delivery falls by 0.107 points and knowledge of transmission through breastfeeding by 0.080 points. The four-item maternal index declines by 6.7 percentage points. All four estimates are statistically significant at the 1 per cent level.

Table 6: Internet Access and Maternal HIV/AIDS Knowledge: Women

	Pregnancy				Delivery	Breastfeeding	Maternal index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment							
Fibre proximity	-0.100***	-0.099***	-0.100***	-0.099***	-0.107***	-0.080***	-0.067***
	(0.028)	(0.028)	(0.029)	(0.029)	(0.029)	(0.023)	(0.014)
Demographic controls							
Age		0.003***	0.003***	0.003***	0.004***	0.001***	0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married		0.028***	0.035***	0.034***	0.043***	0.033***	0.032***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Household size		-0.001	-0.001	-0.000	-0.000	-0.001	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Socioeconomic controls							
Literacy			0.008*	0.009*	0.005	-0.010**	0.008***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Education			0.005***	0.005***	0.011***	0.004***	0.007***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Radio			0.009	0.008	0.007	0.001	0.008***
			(0.006)	(0.006)	(0.005)	(0.005)	(0.003)
Television			-0.006	0.007	0.005	0.001	0.006
			(0.007)	(0.008)	(0.007)	(0.007)	(0.004)
Wealth			0.002	0.005	0.009**	0.002	0.006***
			(0.004)	(0.003)	(0.004)	(0.004)	(0.002)
Geographic controls							
Rural				0.001	-0.005	0.007	-0.002
				(0.015)	(0.014)	(0.014)	(0.008)
Electricity				-0.031***	-0.009	-0.025***	-0.008*
				(0.009)	(0.008)	(0.009)	(0.005)
Constant	1.515***	1.413***	1.355***	1.352***	1.269***	1.399***	0.529***
	(0.009)	(0.012)	(0.017)	(0.030)	(0.030)	(0.030)	(0.016)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	No	No	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	No	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women	Women	Women
Mean dependent variable	1.485	1.485	1.485	1.486	1.481	1.464	0.662
Observations	165111	165111	163810	163785	163777	163777	167868
R^2	0.121	0.124	0.125	0.125	0.121	0.135	0.137

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The dependent variables measure knowledge that HIV/AIDS can be transmitted from mother to child during pregnancy, delivery, and breastfeeding. The maternal index is a four-item measure combining general mother-to-child transmission with the pregnancy, delivery, and breastfeeding indicators. The index allows for partial response, but a strict non-missing index yields consistent results. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The pregnancy coefficient is −0.100 with grid and region-by-year fixed effects, and is stable upon adding demographic, socioeconomic and geographic controls. The final estimate is approximately 6.7 per cent of the estimation-sample mean of 1.486. We use it as the benchmark for the remainder of the natural-experiment analysis.

The three route-specific outcomes establish the breadth of the decline. Their similar magnitudes also reduce the concern that the wording of any single question drives the

results. These outcomes are measured on a three-category scale, with zero indicating an incorrect answer, one a "don't know" response, and two the correct answer. The maternal index averages four binary indicators: the general mother-to-child-transmission, pregnancy, delivery, and breastfeeding. The index ranges from zero to one. Its coefficient therefore corresponds to a 6.7 percentage point decline and provides the most direct interpretation of the treatment effect. The index is also used for several robustness exercises.

The pooled estimates, which include men and women, are smaller but remain negative: pregnancy knowledge falls by 0.047 points on the 0–2 scale and the maternal index by 3.2 percentage points. The estimates are statistically significant at the 5 and 1 per cent levels, respectively. The male-only pregnancy and maternal-index coefficients are positive, at 0.082 and 0.051. However, these coefficients are only suggestive evidence that fibre exposure increased men's maternal-health knowledge as they do not hold up to the various robustness exercises. Appendix B.2 reports the male and pooled estimates, and robustness checks in more detail.

4.6 Mechanisms - Changes in the information environment

The total derivative in equation (12) identifies three channels through which digital access can negatively affect knowledge. First, greater attention to modern media, a_t, can reduce attention to established information sources. Second, declining trust in an accurate traditional source, δ_{Tt}, lowers its effective quality, $Q_{Tt} = \delta_{Tt} q_{Tt}$. Third, greater exposure to modern information can raise Ω_t beyond the point at which the marginal processing cost of additional content exceeds its learning value. The AGIE directly assigns information environments and source substitution; the analysis below examines whether fibre exposure changes source usage, trust, and uncertainty. Because the surveys do not record attention allocations or the source of each respondent's knowledge, these estimates provide indirect mechanism evidence rather than a formal mediation decomposition.

Established health-information channels: The attention constraint in equation (1) allows greater use of modern sources to reduce attention to established sources. Equation (8) shows that such a reallocation lowers knowledge when traditional attention has the higher net marginal return. We cannot observe this reallocation directly. We instead examine whether fibre proximity changes reported exposure to established reproductive-health messages and contact with health providers.

Table 7 shows that fibre proximity reduces reported exposure to family-planning information through radio by 2.8 percentage points and through newspapers or magazines by 1.0 percentage point. Both estimates are statistically significant at the 5 per cent level. These results provide evidence of substitution away from two clearly traditional sources of reproductive-health information.

The television estimate is statistically insignificant. Television is a less clear test of traditional-media substitution than radio or newspapers. It carries scheduled broadcast programming, which could contain family planning information, but can also provide access to streamed and other internet-enabled content. The DHS does not distinguish between these forms of viewing. The television result therefore provides no evidence of reduced exposure through this medium, but it does not necessarily contradict the declines observed for radio and newspapers.

Fibre proximity reduces reported contact with a health or family-planning worker by 4.5 percentage points in the full available sample and by 7.6 points among respondents who answered the HIV module. This decline could reflect reduced demand for in-person advice as online health information becomes easier to obtain. A supply-side possibility is that governments and health providers shifted some information provision towards digital or e-health channels as connectivity improved. The DHS does not measure local service provision, so the regressions cannot distinguish these explanations.

By contrast, fibre proximity has no statistically significant effect on antenatal (pregnancy) visits. With visits capped at 20, the estimates are -0.059 (SE 0.099) in the full mothers sample and -0.083 (SE 0.169) in the HIV-questionnaire sample. The absence of any detectable change in antenatal attendance weighs against a broad contemporaneous reorganisation of health-service provision along fibre corridors, making a change in how women obtained reproductive-health information the more plausible interpretation.

Table 7: Fibre Proximity and Traditional Sources of Health Information

	Family-planning information source			Contact with health providers	
	Radio	Television	Newspaper	Health-worker visit	Antenatal visits
	(1)	(2)	(3)	(4)	(5)
Fibre proximity	-0.028**	0.004	-0.010*	-0.076***	-0.083
	(0.013)	(0.010)	(0.005)	(0.024)	(0.169)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Mothers
Mean dependent variable	0.460	0.304	0.061	0.139	4.257
Observations	189368	189339	189262	164380	79506
R^2	0.203	0.316	0.152	0.167	0.455

Notes: The road-restricted sample retains all observations in the applicable fibre-proximity treatment zone and restricts comparison observations outside that zone to locations within 10 kilometres of a main road. Columns (1)–(3) indicate whether respondents reported receiving family-planning information through radio, television, or newspapers. Column (4) indicates a health-worker visit, while column (5) reports the number of antenatal visits, with counts of 20 or more coded as 20. Each specification includes the controls defined for the corresponding outcome, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fibre exposure is therefore accompanied by less family-planning information through radio and newspapers and fewer contacts with health workers, but no detectable decline in antenatal attendance. The evidence suggests selective substitution away from established information channels rather than a general withdrawal from formal health care.

Trust in official information: Trust links institutional confidence to the model's effective-quality term. When public-health messages carried by official media and health systems are accurate, $q_{Tt} > 0$, so a fall in δ_{Tt} reduces Q_{Tt} and the knowledge gained from them even if their content is unchanged. One route is that online access makes government misconduct more visible. Guriev et al. (2021) show that 3G expansion reduces government approval and makes documented corruption more visible, with corruption revelations translating more strongly into perceived corruption in covered areas.

We find consistent results: fibre proximity lowers trust in the president, ruling party, local government, parliament, and electoral commission, with coefficients between -0.071 and -0.084 points.

Table 8: Fibre Proximity and Trust in Political Institutions

	Afrobarometer				
	President	Ruling party	Local government	Parliament	Electoral commission
	(1)	(2)	(3)	(4)	(5)
Fibre proximity	-0.0726***	-0.0709***	-0.0842***	-0.0739***	-0.0790***
	(-3.71)	(-3.56)	(-4.66)	(-4.08)	(-4.19)
Country-region FE	Yes	Yes	Yes	Yes	Yes
Country-by-year FE	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	1.820	1.486	1.590	1.556	1.612
Observations	43,121	39,904	42,107	40,521	39,654
R-squared	0.158	0.132	0.130	0.133	0.127

Notes: The dependent variables measure respondent trust in the president, ruling party, local government, parliament, and electoral commission. "Fibre proximity" is the treatment indicator derived from the mapped fibre network. All columns include the Afrobarometer control set, country-qualified region fixed effects, and country-by-survey-year fixed effects. Standard errors are clustered by country-qualified town/village. t statistics are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The decline in trust extends beyond the public institutions themselves and to the confidence of their media. Trust in public media falls by 0.049 points and perceived corruption in public media rises by 0.044 points, while the corresponding estimates for private and social media are smaller and statistically insignificant. The differential movement is consistent with a decline in the subjective quality assigned to public information sources, as a consequence of greater scepticism of the providers of this information.

Table 9: Fibre Proximity, Media Trust, and Perceived Media Corruption

	Publicly owned media		Privately owned media		
	Trust in public media	Corruption in public media	Trust in social media	Trust in private media	Corruption in private media
	(1)	(2)	(3)	(4)	(5)
Fibre proximity	-0.0487**	0.0444**	-0.0142	-0.0150	0.0272
	(-2.07)	(2.27)	(-0.56)	(-0.63)	(1.17)
Country-region FE	Yes	Yes	Yes	Yes	Yes
Country-by-year FE	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	1.530	1.225	1.290	1.531	1.224
Observations	8,907	8,524	8,139	8,873	6,262
R^2	0.183	0.042	0.114	0.127	0.051

Notes: The dependent variables measure trust in public media, perceived corruption in public media, trust in social media, trust in private media, and perceived corruption in private media. These questions are only available in the Ghana Afrobarometer surveys for rounds 7–10. "Fibre proximity" is the treatment indicator derived from the mapped fibre network. All columns include the Afrobarometer control set, country-qualified region fixed effects, and country-by-survey-year fixed effects. Standard errors are clustered by country-qualified town/village. t statistics are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Pre-rollout trust in government broadcasting, independent broadcasting, government newspapers, and independent newspapers is similar across eventual treatment groups. This survey question was removed in round 4, so a full pre trend analysis is not possible. Though, these results indicate none of the four differences are statistically significant. The conclusion is unchanged when the comparison is restricted to Ghana. These estimates make pre-existing differences in media scepticism a less plausible account of the post-rollout results.

Table 10: Pre-Rollout Trust in Government and Independent Media

	Publicly owned media		Privately owned media	
	Government broadcasting	Government newspapers	Independent broadcasting	Independent newspapers
	(1)	(2)	(3)	(4)
Ever treated	-0.0158	0.0279	-0.00969	-0.00706
	(-0.52)	(0.71)	(-0.28)	(-0.17)
Country-region FE	Yes	Yes	Yes	Yes
Country-by-year FE	Yes	Yes	Yes	Yes
Mean dependent variable	2.199	2.000	2.142	1.898
Observations	4,574	3,454	4,158	3,399
R-squared	0.299	0.286	0.167	0.185

Notes: The table uses pre-rollout Afrobarometer rounds in which respondents were asked about trust in government and independent media. These questions are available in Ghana rounds 2 and 3, Senegal rounds 2 and 3, and Mali round 3. The estimates test whether locations eventually exposed to fibre infrastructure differed in media trust before rollout. None of the four differences is statistically significant; the same conclusion holds for Ghana-only. All columns include the Afrobarometer control set, country-qualified region fixed effects, and country-by-survey-year fixed effects. Standard errors are clustered by country-qualified town/village. t statistics are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Information overload and uncertainty: Information overload arises when the information to be processed grows faster than the receiver's capacity to screen it (Eppler

and Mengis, 2004). This channel complements the earlier source-substitution mechanism because internet access changes both the composition of information received and the amount of material that must be screened. In the model, a rise in modern information volume, $\bar{V}_{Mt}$, changes effective information, I_t, and increases raw exposure, Ω_t. Before local fibre expansion, access to modern information is limited, so the product $a_{Mt}\bar{V}_{Mt}$ is small rather than necessarily zero. Improved infrastructure expands both the online content available and the opportunity to allocate attention to it. Once the inequality in equation (10) is reversed, the processing cost of additional content dominates its contribution to learning. Importantly, this mechanism does not require online claims to be false. Accurate and partially accurate claims also impose a cognitive burden when respondents must assess their credibility, relevance, and application to their own circumstances.

Afrobarometer records a broad expansion in news use after fibre exposure. On the five-point frequency scale, television news rises by 0.230 points, internet news by 0.121 points, and social-media news by 0.127 points. General radio news changes little. Fibre exposure therefore increases the use of online sources alongside consumption through several media, a pattern consistent with an expanded information set (Ω_t). The results are stable when omitting device ownership controls (Appendix Table 33)

Table 11: Fibre Proximity and News Consumption

	Radio	Television	Internet	Social media
	(1)	(2)	(3)	(4)
Fibre proximity	0.0112	0.230***	0.121***	0.127***
	(0.59)	(7.21)	(6.09)	(5.22)
Country-region FE	Yes	Yes	Yes	Yes
Country-by-year FE	Yes	Yes	Yes	Yes
Mean dependent variable	2.995	2.083	0.832	1.033
Observations	42,282	42,200	33,927	27,251
R-squared	0.250	0.362	0.408	0.413

Notes: The dependent variables measure how often respondents use radio, television, the internet, and social media as sources of news. Each variable is scored from zero to four: never, less than once a month, a few times a month, a few times a week, and every day. "Fibre proximity" is the treatment indicator derived from the mapped fibre network. All columns include the Afrobarometer control set, country-qualified region fixed effects, and country-by-survey-year fixed effects. Standard errors are clustered by country-qualified town/village. t statistics are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fibre proximity also increases uncertainty in the underlying responses. "Don't know" answers rise by 3.0 percentage points for pregnancy, 3.5 points for delivery, and 3.7 points for breastfeeding; all three estimates are significant at the 1 per cent level. Incorrect answers also rise by 3.5, 3.6, and 2.2 percentage points, while correct answers fall by 6.5, 7.1, and 5.8 points, respectively. The increase in "don't know" responses is consistent with greater uncertainty, but it is not clear whether this is driven by the volume or overall quality of information. It may also reflect decreased attention to established sources

following increased digital use. We therefore treat overload as one possible explanation for uncertainty.

Table 12: Fibre Proximity and Responses About Mother-to-Child Transmission

	Incorrect	Don't know	Correct
	(1)	(2)	(3)
Pregnancy	0.035**	0.030***	−0.065***
	(0.014)	(0.009)	(0.016)
Delivery	0.036**	0.035***	−0.071***
	(0.014)	(0.010)	(0.016)
Breastfeeding	0.022**	0.037***	−0.058***
	(0.011)	(0.011)	(0.014)

Notes: Each cell reports the fibre-proximity coefficient from a separate regression of an indicator for the response and transmission stage shown. All regressions include grid and region × year fixed effects. Observations range from 163,777 to 163,785. Dependent-variable means are 0.194–0.203 (incorrect), 0.111–0.132 (don't know), and 0.667–0.687 (correct); R^2 ranges from 0.114 to 0.138. Grid-clustered standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The two datasets provide complementary evidence on the information environment. In the DHS, fibre exposure is accompanied by less family-planning information through radio and newspapers, fewer contacts with health workers, and more "don't know" responses. Afrobarometer records greater use of online news alongside lower trust in public institutions and public media. This joint movement is consistent with women receiving less information through established health channels while facing a larger and more difficult information set to screen.

4.7 Identification and robustness

The checks below focus on the two features of the design most likely to affect the benchmark estimate: the evolution of knowledge around fibre connection and the staggered timing of the rollout.

We next compare conventional TWFE estimates with the stacked design described in Section 4.3. This comparison addresses the possibility that previously treated cells serve as controls for later treatment cohorts and that treatment effects vary across cohorts or exposure lengths.

The event-study estimates provide the first check on the timing of the decline. Figure 4 plots conventional TWFE and stacked estimates for pregnancy-transmission knowledge using the road-restricted sample. Event time −1 is normalised to zero, and the two specifications produce broadly similar pre- and post-treatment paths. In both panels, the earliest pooled lead is positive and statistically significant. Because the ≤ -6 bin

combines all observations recorded at least six years before exposure, it indicates an early historical difference relative to the omitted period rather than a coefficient for one particular pre-treatment year.

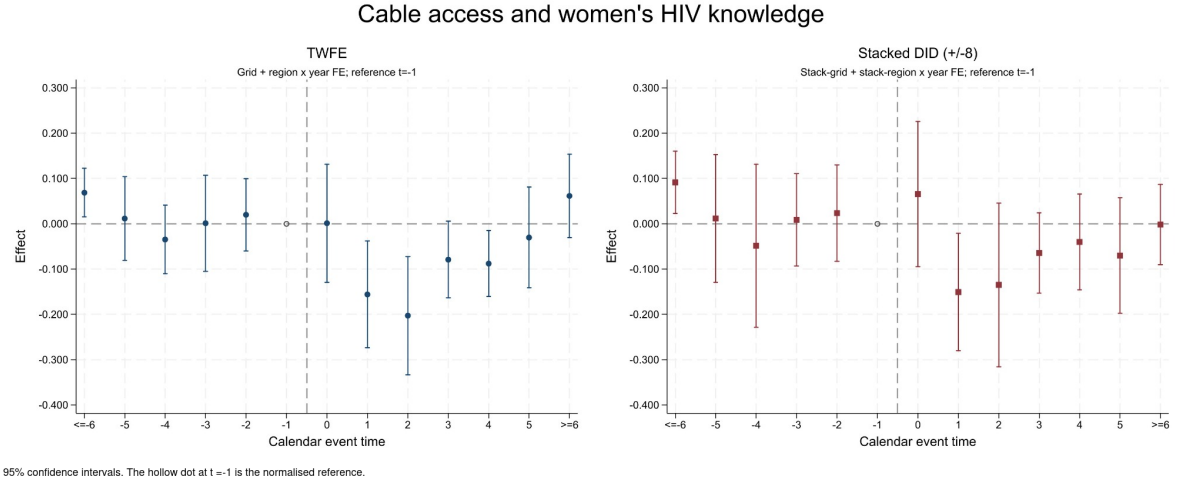


Figure 4: TWFE and stacked event-study estimates for women's pregnancy-transmission knowledge

The pre-treatment path is considerably more stable during the five years immediately preceding connection. The coefficients from event times -5 through -2 are individually statistically insignificant and jointly indistinguishable from zero under both TWFE ($p = 0.815$) and the stacked specification ($p = 0.973$). Including the pooled earliest lead, the full pre-treatment test rejects under TWFE ($p = 0.035$) but not under the stacked specification ($p = 0.173$). The recent pre-treatment coefficients therefore provide no evidence of differential movement before connection, although the earliest pooled lead cautions against claiming parallel trends over the entire available pre-period.

Following treatment, the estimates turn negative. The coefficients at event times $+1$ through $+5$ average -0.111 under TWFE (SE 0.030; $p < 0.001$) and -0.092 under the stacked specification (SE 0.031; $p = 0.003$). The decline is therefore somewhat smaller under the preferred stacked design but remains statistically significant. At longer exposure horizons, the pooled coefficients move towards zero. The coefficient for six or more years after treatment is (0.061) under TWFE (SE 0.047; p=0.190) and (-0.002) under the stacked specification (SE 0.045; (p=0.970)). An early-to-late difference remains statistically significant when adoption-year cohorts and their aggregation weights are held fixed (Appendix Table 31); changes in cohort shares alone therefore do not account for that comparison.

One possible interpretation of this attenuation is that users gradually adapt to the expanded information environment. Digital literacy and scepticism towards low-quality material may improve with experience, reducing the processing costs associated with a larger information set. The supply of reliable online health information, including information provided by governments and public-health organisations, may also have

improved over time. Either development could weaken the initial effects of information overload.

Table 13 compares estimates from the conventional specification with those from the stacked design. In the road-restricted sample, conventional TWFE produces a pregnancy coefficient of −0.099 and a maternal-index coefficient of −0.067. The corresponding stacked estimates are −0.074 and −0.059, respectively, and both remain statistically significant at the 1 per cent level. The negative estimates therefore do not appear to be generated solely by comparisons in which previously treated cells serve as controls. Taken together, the stable pre-treatment path over the five years preceding connection and the persistence of the main coefficients under the preferred stacked design strengthen the evidence from the natural experiment, while the earliest pooled lead warrants continued caution about longer-run historical comparability.

Table 13: Conventional TWFE and Stacked Difference-in-Differences Estimates

	Pregnancy knowledge			Maternal index		
	All controls	Road restricted	Stacked DiD	All controls	Road restricted	Stacked DiD
	(1)	(2)	(3)	(4)	(5)	(6)
Conventional TWFE						
Fibre proximity	-0.103***	-0.099***		-0.072***	-0.067***	
	(0.026)	(0.029)		(0.013)	(0.014)	
Stacked DiD						
Fibre proximity			-0.074***			-0.059***
			(0.027)			(0.014)
Grid fixed effects	Yes	Yes	Stack-grid	Yes	Yes	Stack-grid
Region × year fixed effects	Yes	Yes	Stack-region x year	Yes	Yes	Stack-region x year
Stacked DiD	No	No	Yes	No	No	Yes
Sample	Women: all	Women: road restricted	Women: road restricted	Women: all	Women: road restricted	Women: road restricted
Mean dependent variable	1.483	1.486	1.476	0.654	0.662	0.663
Observations	196292	163785	379535	201733	167868	384645
R^2	0.130	0.125	0.131	0.149	0.137	0.149

Notes: Columns (1), (2), (4), and (5) report conventional TWFE estimates. Columns (3) and (6) report stacked difference-in-differences estimates on the road-restricted sample. Stacked observation counts include repeated appearances across cohort stacks. Fixed effects are reported in the table. Standard errors, clustered at the relevant grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The next set of checks examines whether the results depend on how fibre exposure is assigned geographically. The binary treatment measure approximates local serviceability around the fibre network. Uncertainty about the relevant service radius, together with the displacement of DHS coordinates, can misclassify observations near a treatment boundary. The benchmark rule assigns exposure within 15 kilometres of a fibre node or 1 kilometre of a cable. Table 14 compares this definition with a 15-kilometre node-only treatment and a more conservative rule using a 10-kilometre node radius and 1-kilometre cable radius. The three pregnancy estimates are −0.099, −0.099, and −0.077, and each is statistically significant at the 1 per cent level.

A binary rule also assumes that exposure changes sharply at the treatment boundary, although serviceability is likely to decline more gradually with distance from the network.

We therefore replace the indicator with inverse-distance measures. Using a continuous measure removes the sharp treatment boundary. Any GPS displacement changes exposure along a continuous scale, rather than reclassifying observations as treated or untreated. The measure is calculated as $Inv_i = 1/(1 + Dist_i)$, so it equals one at the cable and approaches zero as distance increases; an alternative measure incorporates proximity to both cables and nodes. The cable-distance estimates are −0.074 for pregnancy knowledge and −0.045 for the maternal index, both significant at the 10 per cent level. The cable-and-node estimates are −0.124 and −0.075, both significant at the 1 per cent level. The larger cable-and-node estimates are consistent with network nodes capturing local distribution capacity that cable distance alone omits. Across these measures, the estimates remain negative, indicating limited sensitivity to the precise spatial rule used to assign exposure.

Table 14: Robustness to Alternative Treatment Definitions

	Alternative treatment radii			Inverse-distance specifications			
	15 km node + 1 km cable	15 km node only	10 km node + 1 km cable	Cable distance		Cable-node distance	
	Pregnancy	Pregnancy	Pregnancy	Pregnancy	Maternal index	Pregnancy	Maternal index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fibre proximity	-0.099***	-0.099***	-0.077***				
	(0.029)	(0.030)	(0.022)				
Inverse distance				-0.074*	-0.045*	-0.124***	-0.075***
				(0.045)	(0.023)	(0.035)	(0.016)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women	Women	Women
Mean dependent variable	1.486	1.485	1.485	1.486	0.662	1.486	0.662
Observations	163785	163579	161800	163785	167868	163785	167868
R^2	0.125	0.125	0.125	0.125	0.137	0.125	0.137

Notes: Columns (1)–(3) report alternative binary treatment definitions. Column (1) uses the preferred definition: within 15 kilometres of a network node or within 1 kilometre of a fibre cable. Column (2) uses a 15-kilometre cable-only treatment. Column (3) uses a 10-kilometre node radius and a 1-kilometre cable radius. Columns (4)–(7) replace the binary treatment with inverse-distance measures scaled between zero and one. Columns (4) and (5) use inverse distance to fibre cables; columns (6) and (7) use inverse distance to cables and nodes. The four-item maternal HIV/AIDS knowledge index combines the general mother-to-child transmission item with pregnancy, delivery, and breastfeeding transmission knowledge. All columns include the main controls, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Road placement creates a related spatial concern. Fibre cables frequently follow main roads because existing infrastructure lowers construction costs. Road proximity may also predict urban growth, access to health services, and exposure to traditional media. Table 15 reports the unrestricted estimate in column (1) as a reference and the road-restricted benchmark in column (2). Columns (3)–(5) instead restrict distance from the fibre network to 50, 25, and 20 kilometres. The pregnancy coefficients range from −0.079 to −0.112 and are statistically significant at the 1 per cent level throughout. The estimate is smallest in magnitude in the 20-kilometre sample.

Table 15: Road Proximity and Distance-Restricted Robustness

	Unrestricted	Near road	Within 50 km	Within 25 km	Within 20 km
	(1)	(2)	(3)	(4)	(5)
Fibre proximity	-0.103***	-0.099***	-0.112***	-0.091***	-0.079***
	(0.026)	(0.029)	(0.027)	(0.029)	(0.030)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women
Mean dependent variable	1.483	1.486	1.484	1.481	1.481
Observations	196292	163785	186973	162469	153652
R^2	0.130	0.125	0.128	0.125	0.125

Notes: The outcome is pregnancy-transmission knowledge. Column (1) uses the full women sample. Column (2) retains all fibre-proximate observations and restricts the comparison group to locations within 10 kilometres of a main road. Columns (3)–(5) retain observations within 50, 25, and 20 kilometres of the nearest fibre cable in operation in the survey year, respectively. These restrictions replace the road restriction in column (2). Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The remaining exercises examine specification choices, sample composition, and falsification outcomes. Appendix Table 34 varies the spatial scale of the fixed effects, the time fixed effects, and the clustering convention. The pregnancy coefficient ranges from −0.044 with ADM2 (district) fixed effects to −0.101 with hexagonal fixed effects; six estimates are significant at the 1 per cent level, while the survey-year specification is significant at the 5 per cent level. The country-by-year specification produces an estimate of −0.092, close to the preferred region-by-year estimate of −0.099; using survey-year fixed effects gives −0.064.

Major cities may receive fibre earlier and differ from other locations in health services, media markets, migration, and economic growth. Appendix Table 37 sequentially excludes the capital, the two largest cities, and the three largest cities from the sample. The city-exclusion estimates range from −0.077 to −0.083 and remain significant at the 1 per cent level. Appendix Table 36 then excludes each of the six countries in turn. The pregnancy estimates remain highly significant and range from −0.056 to −0.128, and the maternal-index estimates from −0.053 to −0.078.

Appendix Table 38 reports falsification estimates for respondent composition, household gender relations, and broad local development. All seven estimates are small and statistically insignificant, reducing the concern that fibre rollout coincided with general changes in sample composition or local public goods.

A separate concern is that the increase in "don't know" responses reflects interview fatigue rather than greater uncertainty. Adding interview duration to the uncertainty regressions leaves the estimates similar (Table 39).

Across the timing, treatment-assignment, and route-placement checks, the estimated effects remain negative. The preferred stacked estimates are smaller than their conventional TWFE counterparts but remain statistically significant. The five-year pre-treatment path

shows no differential movement, and the main estimates persist across fixed effects, clustering conventions, sample restrictions, and spatial treatment definitions.

4.8 Heterogeneity

The heterogeneity analysis returns to the exposure, screening, and maternal-health relevance margins highlighted by the model. Table 16 reports interaction coefficients and implied treatment effects by literacy, television ownership, and wealth. The estimated effect is −0.075 among women unable to read a sentence, −0.141 among women who can read part of a sentence, and −0.118 among women who can read a full sentence. The no-literacy estimate is statistically significant at the 5 per cent level, while the partial- and full-literacy estimates are significant at the 1 per cent level. The partial- and full-literacy interaction coefficients relative to no literacy are also negative and significant, although the table does not test equality between the two literate groups.

Women without literacy may have less direct exposure to written online content, which could explain their smaller estimated effect. The difference between partially and fully literate women aligns closely to the processing and trust margins in the model. Both groups can access written online information, but partially literate users may have a higher λ_i, so a given expansion in Ω_t generates a larger overload cost. They may also place greater subjective weight, δ_{Mt}, on unreliable or misleading online sources, magnifying the effect of content for which $q_{Mt} < 0$. Fully literate women remain highly exposed, but stronger processing or source-evaluation skills could provide some protection. These channels are interpretations of the point-estimate pattern rather than separately identified mechanisms.

Media access and wealth show a similar pattern of treatment effect heterogeneity. The implied effect is −0.116 among women in households with a television and −0.081 among those without one; the estimates are statistically significant at the 1 and 5 per cent levels, respectively, but the interaction between fibre proximity and television ownership is statistically insignificant.

The implied effects are −0.006 and −0.021 in the bottom two wealth quintiles and are statistically insignificant. They become −0.088, −0.133, and −0.146 in the third, fourth, and fifth quintiles, respectively, and are significant at the 1 per cent level. Women in better-off households may be more likely to use newly available internet services because they can afford devices and data. The cash-income comparisons in Table 5 support this interpretation. Covariate checks find no statistically detectable change in television ownership or mean literacy, although some wealth-quintile shares change (Appendix Table **??**). We therefore interpret the subgroup estimates as differences by characteristics observed at interview.

Table 16: Heterogeneous Effects by Literacy, Media Access, and Wealth

	Interaction specification			Implied effect
	Literacy	Television ownership	Wealth	Subgroup total
	(1)	(2)	(3)	
Fibre proximity × reference group	-0.075**	-0.081**	-0.006	
	(0.029)	(0.035)	(0.034)	
Fibre proximity × partial literacy	-0.066***			-0.141***
	(0.017)			(0.034)
Fibre proximity × full literacy	-0.043***			-0.118***
	(0.015)			(0.029)
Fibre proximity × owns television		-0.036		-0.116***
		(0.022)		(0.027)
Fibre proximity × wealth quintile 2			-0.015	-0.021
			(0.022)	(0.030)
Fibre proximity × wealth quintile 3			-0.082***	-0.088***
			(0.027)	(0.030)
Fibre proximity × wealth quintile 4			-0.127***	-0.133***
			(0.031)	(0.034)
Fibre proximity × wealth quintile 5			-0.141***	-0.146***
			(0.036)	(0.033)
Grid fixed effects	Yes	Yes	Yes	
Region × year fixed effects	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	
Sample	Women	Women	Women	
Mean dependent variable	1.486	1.486	1.486	
Observations	163785	163785	163785	
R^2	0.125	0.125	0.125	

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The dependent variable is pregnancy-transmission knowledge. The reference groups are women with no literacy, women without a television, and women in the poorest wealth quintile. Columns (1)–(3) report interaction coefficients relative to those groups; the final column reports the corresponding total subgroup effects calculated after estimation. A dot indicates that a margin was not calculated. All specifications include the main controls, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Another useful comparison is between individuals with varying relevance for maternal health information. Table 17 reports estimates by age and motherhood. The estimate is -0.104 for women aged 40 or younger and -0.067 for older women; the interaction difference is significant at the 5 per cent level. Mothers have an implied effect of -0.106, compared with -0.087 among women who are not mothers, and the difference is significant at the 10 per cent level. Younger women and mothers are more likely to encounter maternal-health information because it is more directly relevant to them. Age also captures differences in digital adoption, so both relevance and exposure may contribute to the variation.

Table 17: Heterogeneous Effects by Maternal Health Salience

	Pregnancy knowledge		Implied effect
	Age 40 or below	Motherhood	Subgroup total
	(1)	(2)	
Fibre proximity × reference group	-0.067**	-0.087***	
	(0.032)	(0.028)	
Fibre proximity × age 40 or below	-0.036**		-0.104***
	(0.015)		(0.029)
Fibre proximity × mother		-0.019*	-0.106***
		(0.011)	(0.030)
Grid fixed effects	Yes	Yes	
Region × year fixed effects	Yes	Yes	
Controls	Yes	Yes	
Sample	Women	Women	
Mean dependent variable	1.486	1.486	
Observations	163785	163785	
R^2	0.125	0.126	

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The dependent variable is pregnancy-transmission knowledge. Column (1) interacts fibre proximity with an indicator for women aged 40 or below; the reference group is women over age 40. Column (2) interacts fibre proximity with motherhood; the reference group is women who are not mothers. The final column reports the corresponding total subgroup effects calculated after estimation. A dot indicates that a margin was not calculated. All specifications include the main controls, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 40 reports estimates by antenatal-care use for the pregnancy covered by the antenatal-care questions. The effects are statistically insignificant for women with no visits or 1–3 visits. They are −0.093, −0.137, and −0.212 for women reporting 4–6, 7–9, and at least 10 visits, respectively, and are significant at the 1 per cent level. Because antenatal-care use may itself respond to fibre exposure, this remains supportive.

The treatment heterogeneity is consistent with differences in exposure, screening, and the relevance of maternal-health information. Among women, the treatment effect becomes more negative in groups who are more likely to engage with the internet and with maternal health information.

5 Conclusion

This paper asks whether digital access can reduce verified knowledge when information becomes easier to supply than to screen. The Agentic General Intelligence Experiment shows how source substitution can reduce knowledge within a controlled model environment. When compared with receiving no new information, even the noisy online feed improves

knowledge of maternal HIV transmission, although by less than high-quality traditional content. After a common period of exposure to traditional media, however, agents who switch to the noisy online feed score 10.7 percentage points lower than those who continue receiving traditional content. Knowledge falls among the agents who switch and continues to rise among those who remain with traditional media.

Exploiting the staggered rollout of fibre-optic infrastructure across six West African countries, we find that fibre proximity increases internet use but lowers women's knowledge of mother-to-child HIV transmission. The decline covers transmission during pregnancy, delivery, and breastfeeding. Estimates from the preferred stacked design are smaller than conventional TWFE estimates but remain negative and statistically significant.

Fibre expansion also changes the sources from which women receive reproductive-health information. Exposure to family-planning messages through radio and newspapers declines, as does contact with health workers, while the estimates for television and antenatal visits provide little evidence of a corresponding change. Online news use and uncertainty increase, and confidence in public institutions and public media falls.

The two designs support an account in which expanded digital access displaces some established sources of health guidance. The AGIE shows why such displacement could matter even when online content contains useful information: a noisy feed can replace a reliable source that would otherwise have continued reinforcing knowledge. The surveys provide supporting evidence of changes in source use and trust after fibre expansion. Scenario V imposes complete substitution within a language model. Information overload remains a possible additional channel: the experiment holds feed length fixed, and the field measures do not separately identify its contribution.

The policy implication is to preserve access to verified guidance as media use changes. Repeated health messages, verified sources, in-person advice, and digital-literacy training can complement internet expansion. Verified maternal-health guidance could continue through radio and health workers while also being adapted to the online platforms people increasingly use. The relevant measure of digital progress is ultimately whether people acquire reliable knowledge from the larger information set available to them.

References

Abeler, J., Huffman, D., and Raymond, C. (2025). Incentive complexity, bounded rationality, and effort provision. *American Economic Review*, 115(12):4404–4437.

Acemoglu, D., Ozdaglar, A., and Siderius, J. (2024). A model of online misinformation. *The Review of Economic Studies*, 91(6):3117–3150.

Africa Bandwidth Maps (2026). African terrestrial fibre network maps. https://www.africabandwidthmaps.com/. Accessed 2 September 2026.

AfTerFibre (2026). African terrestrial fibre network database. https://afterfibre.opentelecomdata.org/. Accessed 2 September 2026.

Aher, G. V., Arriaga, R. I., and Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 337–371. PMLR.

Amaral-Garcia, S., Nardotto, M., Propper, C., and Valletti, T. (2022). Mums go online: Is the internet changing the demand for health care? *The Review of Economics and Statistics*, 104(6):1157–1173.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Betts, K. R., Aikin, K. J., Kelly, B. J., Johnson, M., Parvanta, S., Southwell, B. G., Mack, N., Tzeng, J., and Cameron, L. (2019). Taking repeated exposure into account: An experimental study of direct-to-consumer prescription drug television ad effects. *Journal of Health Communication*, 24(5):503–511.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2016). Competition for attention. *The Review of Economic Studies*, 83(2):481–513.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Caplin, A., Dean, M., and Martin, D. (2011). Search and satisficing. *American Economic Review*, 101(7):2899–2922.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., and Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, 19(11):1095–1102.

Chen, H. and Suen, W. (2023). Competition for attention and news quality. *American Economic Journal: Microeconomics*, 15(3):1–32.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Dean, M. and Neligh, N. (2023). Experimental tests of rational inattention. *Journal of Political Economy*, 131(12):3415–3461.

Donati, D. (2023). Mobile internet access and political outcomes: Evidence from south africa. *Journal of Development Economics*, 162:103073.

Eppler, M. J. and Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5):325–344.

Galperti, S. and Trevino, I. (2020). Coordination motives and competition for attention in information markets. *Journal of Economic Theory*, 188:105039.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

Guan, M., Oprea, R., and Yuksel, S. (2023). Too much information. Working paper, July 15, 2023.

Guriev, S., Melnikov, N., and Zhuravskaya, E. (2021). 3G internet and confidence in government. *The Quarterly Journal of Economics*, 136(4):2533–2613.

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–1079.

Horton, J. J., Filippas, A., and Manning, B. S. (2026). Large language models as simulated economic agents: What can we learn from homo silicus? Working Paper 31122, National Bureau of Economic Research. Revised February 2026.

Iacoella, F., Gassmann, F., and Tirivayi, N. (2022a). Impact of mobile phones on HIV public stigma: A cross-sectional and pseudo-panel analysis from ghana. *BMJ Open*, 12(11):e062594.

Iacoella, F., Gassmann, F., and Tirivayi, N. (2022b). Which communication technology is effective for promoting reproductive health? television, radio, and mobile phones in sub-saharan africa. *PLoS ONE*, 17(8):e0272501.

Iacoella, F. and Tirivayi, N. (2024). Mobile phones and HIV testing: Multicountry instrumental variable analysis from sub-saharan africa. *Journal of Medical Internet Research*, 26:e48794.

Infrapedia (2026). Global internet infrastructure map. https://www.infrapedia.com/. Accessed 2 September 2026.

ITU (2001). Mobile technology.

ITU (2026). Broadband map. https://bbmaps.itu.int/bbmaps/. Accessed 2 September 2026.

Iyamu, I., Oladele, E. A., Eboreime, E., and Karim, M. E. (2021). Is regular access to internet services associated with comprehensive correct HIV/AIDS knowledge among people aged 15–49 years in nigeria? findings from the 2018 demographic health survey. *Journal of Consumer Health on the Internet*, 25(3):242–260.

Jacoby, J., Speller, D. E., and Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, 11(1):63–69.

Kergall, P. and Guiffard, J.-B. (2024). Impact of broadband internet on preventive healthcare behaviors in senegal. *Economie et Statistique / Economics and Statistics*, (542):99–123.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–70.

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., et al. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, 7(1):74–101.

Lucas, R. E. J. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Mensah, J. T., Tafere, K., and Abay, K. A. (2022). Saving lives through technology: Mobile phones and infant mortality. Policy Research Working Paper 9978, World Bank.

Mo, D., Luo, R., Liu, C., Zhang, H., Zhang, L., Medina, A., and Rozelle, S. (2014). Text messaging and its impacts on the health and education of the poor: Evidence from a field experiment in rural China. *World Development*, 64:766–780.

Mutumba, M. (2022). Mass media influences on family planning knowledge, attitudes and method choice among sexually active men in sub-saharan africa. *PLoS ONE*, 17(1):e0261068.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., and Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5):812–822.

Serra-Garcia, M. (2026). The attention-information trade-off. *American Economic Review*, 116(5):1579–1610.

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., and Perez, E. (2024). Towards understanding sycophancy in language models. In *International Conference on Learning Representations*.

Simon, H. A. (1971). Designing organizations for an information-rich world. In Greenberger, M., editor, *Computers, Communications, and the Public Interest*, pages 37–72. Johns Hopkins Press, Baltimore, MD.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Suarez-Lledo, V. and Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research*, 23(1):e17187.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International Economic Review*, 6(1):18–31.

Wing, C., Freedman, S. M., and Hollingsworth, A. (2024). Stacked difference-in-differences. Working Paper 32054, National Bureau of Economic Research.

World Bank (2024). Digital progress and trends report: Interactive charts – digital adoption.

World Bank (2026). Individuals using the internet (% of population). World Development Indicators, indicator IT.NET.USER.ZS. Data sourced from the International Telecommunication Union, World Telecommunication/ICT Indicators Database. Accessed 10 August 2026.

World Health Organization (2024a). Mother-to-child transmission of hiv.

World Health Organization (2024b). Pregnant and lactating women and girls.

A Agentic-Experiment Appendix

This appendix records the information treatments and agent-facing prompts, then reports question-level, iteration, and sampling-temperature results omitted from the main experimental sequence.

A.1 Information feeds

Table 18 reproduces the 225-item master corpus from which the experimental feeds are sampled. The corpus contains 75 accurate items and 50 items in each of the confusing, misleading, and irrelevant categories. Scenario-specific sampling changes the shares drawn from these four pools while holding the feed length fixed at 50 items. The categories describe message content and presentation. Confusing items can therefore contain accurate guidance alongside ambiguity, or sensationalist framing.

Table 18: Information-Item Corpus Used in the Experiment

Category	Information sentences available for sampling
Correct / accurate	Pregnant? Ask for an HIV test at your first antenatal visit. Starting care early protects both mother and baby.
	I tested during pregnancy even though I felt well. The midwife said knowing early gave us more ways to protect my baby.
	More antenatal clinics are offering HIV testing and treatment before labour begins.
	Women living with HIV can give birth to HIV-negative babies when they receive the right care and keep taking treatment.
	Why do nurses ask about HIV before delivery? Because prevention starts during pregnancy and continues after the baby is born.
	HIV care in pregnancy includes treatment, planning for delivery, and follow-up for mother and infant.
	My sister's HIV clinic did not stop helping when she gave birth. They followed her treatment and the baby's feeding and tests.
	A new mother-to-child HIV programme links antenatal care, safe delivery, infant feeding advice, and early testing for babies.
	HIV can pass through certain sexual contact, blood exposure, and from mother to child. It is not spread by ordinary social contact.
	It is not only gay people who can get HIV. Anyone can be exposed through unprotected sex.
	Condoms, HIV testing, and treatment all have a place in preventing new infections.
	Health workers are pairing malaria-bed-net campaigns with separate advice on HIV testing and safer sex because the diseases spread differently.
	You can share food, hug, and live with someone who has HIV. Prevention advice focuses on sex, blood, and treatment.
	Mosquito bites cause malaria, not HIV. Bed nets prevent malaria, while HIV prevention requires different measures.
	I used to worry about sharing cups. The counsellor explained that HIV is not passed through food, cups, hugs, or an ordinary kiss.
	Couples are encouraged to test together rather than trying to judge HIV status from appearance.
	A person may have HIV and feel completely healthy. Only an HIV test can tell you your status.
	Nothing looked or felt wrong when I went for a routine test. Looking healthy is not the same as knowing your HIV status.
	Why waiting for symptoms can delay an HIV diagnosis
	Community testing teams say many visitors arrive without symptoms and simply want to know their status.
	You cannot tell by someone's weight, skin, work, or clothes. A test gives the answer.
	HIV testing is confidential. Ask your local health centre when and where testing is available.
	Testing before marriage or pregnancy is about knowing your status, not accusing anyone of being unfaithful.
	Antiretroviral treatment controls HIV and helps people live long, healthy lives. It must be taken as prescribed.
	My health improved after I started treatment, but I still collect my tablets and attend the clinic. Feeling well does not mean stopping.
	Researchers continue to study an HIV cure while clinics rely on lifelong antiretroviral treatment.
	Undetectable means treatment has reduced the virus to a very low level. It does not mean a person should stop taking medicine.
	HIV treatment has transformed survival, but a generally available cure remains out of reach
	Do not stop HIV treatment because you feel better. Speak to a health professional about side effects or difficulties taking medicine.
	Good HIV care is more than tablets: regular appointments, testing, support, and reliable access to medicine all matter.

Continued on next page

Table 18 – continued from previous page

Category	Information sentences available for sampling
	My neighbour has HIV, works every day, and looks after her family. Her diagnosis does not change how we greet or eat together.
	You cannot catch HIV by sharing a home, workplace, classroom, meal, or handshake.
	Supporting someone to test and stay in care is more useful than blaming them.
	Anti-stigma groups say fear of gossip still prevents some people from seeking an HIV test.
	A healthy appearance tells you nothing certain about HIV, and an HIV diagnosis tells you nothing about a person's character.
	We still hug our brother and share meals with him. His nurse explained that ordinary family contact is safe.
	Living well with HIV: work, family, treatment, and the damage caused by stigma
	Free and confidential HIV counselling and testing are available this week at participating health centres.
	Mobile clinics will visit rural communities to offer antenatal checks, HIV tests, and referrals for treatment.
	If a test is positive, the clinic can explain treatment and the steps that protect partners and babies.
	I was frightened to ask, but the counsellor answered privately and did not judge me.
	People already taking HIV medicine should contact the clinic before their supply runs out.
	Health centres are training midwives to connect pregnant women with HIV treatment and infant follow-up.
	A home remedy may ease a symptom, but HIV testing and treatment require trained health services.
	HIV attacks the immune system, but treatment can keep the virus controlled for many years.
	What three decades of HIV treatment have changed, and what they have not
	HIV and AIDS are related but not identical: untreated HIV can weaken the immune system and lead to AIDS.
	Accurate HIV information protects health. Ask a trained provider when a message online leaves you uncertain.
	I heard many stories before my test. Speaking with a counsellor helped me separate treatment facts from gossip.
	Prevention, testing, and treatment work together. No single poster or remedy replaces all three.
	Women living with HIV should begin or continue antiretroviral treatment during pregnancy and attend follow-up appointments after delivery.
	Early infant HIV testing helps health workers identify infection and begin care quickly when it is needed.
	The clinic tested my baby after birth and arranged another visit later. One early appointment was not the end of the follow-up.
	Partners can support a healthy pregnancy by testing, attending appointments when invited, and helping mothers continue treatment.
	HIV prevention includes condoms, testing, treatment, and medicines that can reduce risk before or soon after some exposures.
	Health services are expanding access to medicines that HIV-negative people can use to lower their chance of acquiring HIV.
	A negative HIV test tells you your own status at the time of testing; it does not reveal your partner's result.
	Razors, needles, and other sharp instruments should not be shared unless they have been properly sterilised.
	No HIV test can detect every new infection immediately. A health worker may recommend another test after a recent exposure.
	Rapid HIV tests can provide results quickly, while a preliminary positive result still requires confirmation.
	My first test was soon after the event I was worried about, so the counsellor asked me to return after the testing window.
	A negative result is not permission to stop prevention; it is a chance to discuss how to remain HIV-negative.
	Starting HIV treatment promptly protects the immune system and lowers the amount of virus in the body.
	When treatment keeps HIV undetectable, the virus is not transmitted sexually, provided treatment and monitoring continue.
	I take my medicine at the same time each day. Missing doses can make it harder to keep the virus controlled.
	Treatment can make HIV a manageable long-term condition, but clinic follow-up remains important even when a person feels healthy.

Continued on next page

Table 18 – continued from previous page

Category	Information sentences available for sampling
	People living with HIV have the same rights to health care, education, work, and family life as everyone else.
	After my diagnosis I expected people to avoid me. The most helpful friends were the ones who listened and treated me normally.
	Community organisations say confidential testing and respectful care help more people seek services without fear of gossip.
	Anyone worried about a recent possible HIV exposure should contact a health service promptly because preventive treatment is time-sensitive.
	Before travelling, collect enough HIV medicine and ask the clinic what to do if an appointment must be changed.
	Health workers are integrating HIV services with antenatal care, tuberculosis screening, and other routine clinic visits.
	HIV weakens the immune system when untreated; AIDS describes the most advanced stage of HIV disease.
	Better treatment has changed what life with HIV can look like, but it has not changed the basic need for testing and prevention.
	Reliable HIV information should explain both what creates risk and what does not, so fear does not replace prevention.
Confusing	The pregnancy advice that has mothers asking whether HIV risk now belongs to the past. Preview: treatment has reduced infant infections, but care still begins before birth.
	User 1: The baby is protected before birth. User 2: Then why do the tablets start during pregnancy? User 3: Ask the clinic because it depends.
	They told my cousin not to panic because treatment makes the risk very low. She came home saying there was no risk at all.
	Most babies of treated mothers remain HIV-negative, although the report gives little detail about when transmission can occur.
	Natural birth or surgery? The answer for mothers living with HIV is not as simple as you think. Preview: treatment history and individual clinical advice matter more than one rule for everyone.
	If the mother takes medicine, can she breastfeed normally? Everyone in this group seems to have received different advice.
	The nurse said to keep taking treatment after delivery. I could not tell whether that was for the mother, the milk, or the baby.
	Infant-feeding guidance for women with HIV changes with treatment access and local conditions, leaving many families unsure what applies to them.
	The everyday habit people fear more than unprotected sex, and what doctors say about it
	User 1: Mosquitoes carry blood. User 2: They carry malaria, not every blood disease. User 3: I still use my net for everything.
	Someone told me an ordinary kiss is safe, but then warned me about tiny cuts that nobody can see.
	Can one bite really pass an infection? Read this before the rainy season. Preview: the article begins with mosquito fears before separating malaria from HIV.
	The programme distributed condoms and bed nets at the same event without clearly separating which disease each item prevents.
	People keep saying HIV is not spread casually, but nobody agrees on what counts as casual.
	I know sharing food is fine. Kissing feels different, though, and the advice online is never simple.
	Protect your family from infection this rainy season: use a bed net, attend testing, and listen for more advice.
	Five changes in the body that could mean more than ordinary tiredness. Preview: each symptom has many possible causes, and an HIV test is the only way to know.
	He looked strong, so the result surprised everyone. Perhaps the test found it before he was really ill.
	User 1: You can see when someone is sick. User 2: Not at first. User 3: Sometimes people just hide it well.
	A testing campaign focuses on people who feel healthy, but its poster also shows familiar images of weight loss and illness.
	Would you recognise HIV if it were standing in front of you?
	No symptoms, no worries? I keep seeing this phrase, but the clinic still says everyone should test.
	Know the signs, know your status, get tested early.
	Scientists report another person free of detectable HIV: is the cure finally here? Preview: the patient received an exceptional cancer treatment that is not available as routine HIV care.
	User 1: Undetectable means gone. User 2: It means controlled. User 3: Either way, the tablets worked.

Continued on next page

Table 18 – continued from previous page

Category	Information sentences available for sampling
	She regained her weight and returned to work after treatment. The family now says she has recovered completely.
	A rare medical case is described as an HIV cure, while the final paragraph says the procedure is not a routine treatment.
	The HIV breakthrough raising hope, and the detail most headlines leave until the end. Preview: the advance may improve treatment but is not a generally available cure.
	My tablets made the virus disappear from the test. I am still trying to understand why the clinic says I must continue.
	Is a disease cured when medicine keeps it away for years, or only when no medicine is needed?
	How close is too close when someone in the family is living with HIV? Preview: ordinary household contact is safe, although the article spends most of its space discussing rare blood exposure.
	I know we should not exclude anyone, but I would still keep separate cups until I understood everything.
	User 1: Treat people normally. User 2: Yes, but being careful cannot hurt. User 3: Careful about what?
	A family says life remained normal after an HIV diagnosis, although relatives quietly changed several household routines.
	She looks healthier than all of us now. Maybe treatment means there is no longer anything to pass on.
	The reason some healthy-looking couples are having a difficult conversation tonight
	Support, caution, and understanding: three things every family affected by HIV needs.
	New HIV services are available. Visit the health centre to learn which tests, medicines, and prevention options are right for you.
	The clinic gave me advice that sounded different from the radio. Maybe both are right for different people.
	What your local clinic may not explain until after an HIV test
	Do you go to the pharmacy, the maternity clinic, or the testing centre first? Nobody answered the same way.
	A new programme promises simpler HIV care, but patients are still waiting for details about eligibility and follow-up.
	Bring your questions about pregnancy, feeding, partners, and treatment to tomorrow's health meeting.
	I heard treatment is free, then someone said the tests around it still cost money. Has anyone gone recently?
	Everything people thought they knew about HIV is changing. Preview: treatment has improved greatly, while the established routes of transmission remain the same.
	The more I read, the less certain I feel. Treatment is better, but does that mean the old advice is wrong?
	User 1: HIV and AIDS are the same. User 2: One leads to the other. User 3: The names do not matter if treatment works.
	New progress against HIV is described as both a turning point and a reminder that prevention is still needed.
	This one HIV fact surprised almost everyone we asked
	Not saying what happened, but everyone should rethink what they believe about HIV. I will explain later.
Misleading	Why more mothers are refusing unnecessary HIV tablets during pregnancy
	The womb keeps the mother's blood away from the baby, so HIV cannot reach a child before birth.
	User 1: Babies only catch HIV after they are born. User 2: Exactly, pregnancy itself protects them.
	Doctors finally admit that infant HIV begins in the delivery room
	A natural birth keeps the baby safe from HIV because there is no surgical blood.
	Wash the newborn immediately after delivery to remove the HIV risk.
	The breast-milk warning that formula companies do not want mothers to question
	Breast milk is made to protect babies, so it cannot carry HIV.
	Rainy-season HIV cases rise again: the mosquito link nobody is discussing
	A mosquito takes one person's blood and puts a little into the next person. That is how HIV moves.
	User 1: My cousin got HIV after many mosquito bites. User 2: That cannot be a coincidence.
	One spray protects your household from malaria, dengue, and HIV-carrying mosquitoes.
	The hidden infection risk in an ordinary kiss
	Do not share cups or kisses with anyone whose HIV status you do not know.
	User 1: Saliva is a body fluid. User 2: Then kissing must carry HIV too.

Continued on next page

Table 18 – continued from previous page

Category	Information sentences available for sampling
	Use mouthwash after kissing to kill any HIV left in the mouth.
	Seven facial signs that reveal a person's HIV status
	You can tell someone is free of HIV when their eyes are clear and their weight is steady.
	User 1: He cannot have HIV; look how strong he is. User 2: True, sick people do not look like that.
	Skip the clinic: this simple mirror test shows whether HIV is weakening your body
	People with HIV always become thin within a few months, so there is no need to test someone who looks healthy.
	Check your tongue and fingernails at home to know your HIV status in minutes.
	Healthy skin and a good appetite are the body's natural HIV-negative result.
	The seven-day HIV cure clinics do not want you to hear about
	Three months of tablets removed the HIV from my blood, so I stopped taking them.
	Clear HIV naturally with this herbal course. No lifelong medicine needed.
	User 1: Undetectable means cured. User 2: Yes, there is nothing left for medicine to treat.
	One kitchen ingredient may eliminate HIV faster than antiretroviral treatment
	Once a person regains weight on HIV medicine, the virus has left the body.
	Stop paying for HIV tablets: cleanse the virus from your body in ten days.
	Fasting for one week starves HIV until it disappears.
	I support people with HIV, but I would never let them cook for the family.
	User 1: Children can catch HIV from a parent's kiss. User 2: Better to keep some distance just in case.
	The household objects people living with HIV should never share
	You can stay friendly with someone who has HIV as long as you do not touch their skin.
	Separate plates and bedding are the safest choice when someone at home has HIV.
	Why an innocent family kiss may not be as safe as experts once said
	No blood test needed: send a photograph and receive your HIV result today.
	The clinic only wants people to keep returning so it can sell more medicine.
	Hospitals hide cheap HIV cures while prescribing tablets for life
	User 1: Testing can put HIV into your blood. User 2: That is why I avoid clinic needles.
	One private massage removes HIV toxins without tests or tablets.
	Traditional tonics protect a baby better than antenatal HIV medicine.
	HIV may be caused by stress rather than infection, new post reveals
	Only people with many sexual partners need to think about HIV.
	User 1: HIV affects only certain communities. User 2: Ordinary families do not need to worry.
	The HIV test result that changes when you wash with salt water
	If both partners look healthy, unprotected sex cannot spread HIV.
	Boost your immunity and become completely protected from HIV exposure.
	Prayer alone removes HIV once a person truly believes they are healed.
Irrelevant	Evening rain is expected along parts of the coast, with clearer weather inland.
	The bus queue was longer than the journey today. I should have left home earlier.
	National team names its squad before this weekend's match
	Saturday's neighbourhood clean-up begins at eight in the morning.
	Buy two mobile-data bundles this week and receive an extra evening allowance.
	Does anyone know why onion prices have risen again at the market?
	Road repairs will divert traffic near the central bus station from Monday.
	The fish-and-rice recipe viewers are trying at home
	User 1: That was never a penalty. User 2: Watch the replay before blaming the referee.
	Parents should complete school registration before the end of the month.
	Festival fabrics are now available in six new colours.
	I finally found a way to keep rice dry through the rainy season.
	Groundnut farmers prepare storage rooms ahead of the next harvest.
	Singer releases a live version of the year's most requested song
	The family gathering has moved from Saturday afternoon to Sunday.
	These beach photographs make me want to travel before the rains begin.
	A local furniture shop offers free delivery on selected tables and chairs.
	The theatre company will open its new production on Friday evening.
	User 1: Add more chilli. User 2: Not if the children are eating too.
	Water service will be interrupted for maintenance between noon and four.
	Why this season's family drama has everyone arguing
	I planted mint in an old cooking pot and it is growing better than expected.
	Open a small savings account with no monthly service charge.
	Local schools publish examination dates for final-year pupils.
	User 1: Take the early bus. User 2: It is faster only when the road is clear.

Continued on next page

Table 18 – continued from previous page

Category	Information sentences available for sampling
	Volunteers are collecting books for the neighbourhood reading room.
	Who has a reliable tailor for a dress needed next week?
	Five inexpensive ways to organise a small kitchen
	Fresh mangoes and oranges available beside the north market entrance.
	Fishing boats returned early as winds strengthened offshore.
	Waste collection on the eastern route will take place one day later this week.
	The new song sounded better live than it did on the radio.
	User 1: The blue fabric is better. User 2: The green one will last longer.
	A lost set of keys was found near the bakery this morning.
	Market traders test a new system for reusable shopping bags
	Discounted sandals are available until stocks run out.
	The power returned just as I had finished cooking in the dark.
	A youth football tournament begins next month with teams from eight districts.
	Residents may collect updated transport cards from the municipal office.
	User 1: The ending of that programme made no sense. User 2: There will be another episode.
	The sewing group meets at the usual place after Friday prayers.
	I tried the grilled-fish seasoning everyone recommended and used far too much lemon.
	Coastal town prepares for its annual music and food festival
	Used refrigerator in working condition. Collection only.
	The price of cooking oil remained unchanged at most surveyed markets.
	Applications for market-stall permits close on Thursday afternoon.
	User 1: Water the plants every day. User 2: Not during the rainy week.
	A basket-weaving demonstration will be held outside the community centre.
	I left my umbrella at home because the morning sky looked completely clear.
	Coach promises changes after another difficult away match

A.2 Agent-facing prompts

Tables 19 and 20 reproduce the model-facing instructions and the respondent characteristics supplied at runtime. The prompt separates the system message, profile, baseline answers, recent-media block, survey questions, and response protocol. Baseline calls import each sampled woman's three DHS answers and elicit the four additional outcomes. Post-exposure calls include the seven recorded answers and the assigned feed, but omit experiment details such as scenario labels, treatment status, hypotheses, and earlier model conversations. The prompt record therefore documents both the information received by the model and the experimental details concealed from it.

Table 19: Prompt Template Used for Synthetic Respondent Interactions

Prompt component	Respondent-facing text or inserted content
System message	Answer as one specific research participant: a woman living in Senegal in the mid-to-late 2010s. Answer as this unique person, using: 1. the respondent profile, 2. her recorded answers from an earlier survey, 3. any recent media material included in this prompt, 4. ordinary prior knowledge this person could plausibly hold. The profile describes the respondent's circumstances and background. Use the complete profile to form a coherent individual perspective and natural manner of responding, without applying fixed rules or stereotypes to any characteristic. Answer according to the knowledge and beliefs this respondent would plausibly hold after considering the profile and any recent media material included below. Do not use live internet access. Do not mention that you are an AI model. Do not invent extra campaigns, facts, media exposure, or experiences that are not in the prompt. The respondent knows that she is participating in a research experiment, but she has not been told her assignment, the study hypothesis, the content of any other condition, or the answers researchers expect. Do not infer or describe these undisclosed features of the study. The earlier answers are this respondent's own starting beliefs. They are not expert statements and are not labelled as correct or incorrect. Continue as the same person while allowing her views to remain unchanged or to change in response to the material she encounters. Treat this survey prompt as self-contained. Do not assume the respondent has seen any earlier study material or completed any earlier study round unless it is explicitly included below. Do not reason as a researcher or describe the study design.
Public profile block	Respondent profile: <JSON object containing the public profile variables listed below>
Baseline stage	Recorded answers from this respondent's earlier DHS interview: <Q1 text> Earlier answer: <Q1 recorded answer> <Q2 text> Earlier answer: <Q2 recorded answer> <Q3 text> Earlier answer: <Q3 recorded answer> Answer the four additional baseline questions as the same respondent. The question block below is supplied for Q4-Q7 at this stage.

Continued on next page

Table 19 – continued from previous page

Prompt component	Respondent-facing text or inserted content
Post-exposure stage	This respondent's recorded answers before the current information environment: <Q1-Q7 text and recorded answers> When applicable, the following material is also inserted: Recent media encountered by the respondent: The material below consists of media items, messages, posts, adverts, headlines, or comments the respondent recently encountered. Do not assume the respondent reads, understands, trusts, or remembers all items equally. Let the profile inform plausible attention and interpretation without treating any characteristic as a fixed rule. Answer according to what this person would plausibly take from the material as a whole, rather than treating it as an answer key. Recent media: <scenario-specific exposure feed>
Question block	The question block below is supplied for Q1-Q7 at this stage. Answer the following questions as the respondent. Q1. Can HIV/AIDS be transmitted from a mother to child during pregnancy? Q2. Can HIV/AIDS be transmitted from a mother to child during delivery? Q3. Can HIV/AIDS be transmitted from a mother to child during breastfeeding? Q4. Can a healthy-looking person have HIV/AIDS? Q5. Can HIV/AIDS be spread by mosquito bites? Q6. At present, can HIV be cured? Q7. Can HIV/AIDS be spread by kissing?
Direct-response protocol	Give only the coded answers. Do not explain or show reasoning. For every question use exactly: Q# final answer: YES/NO/DON'T KNOW Q# final answer: YES/NO/DON'T KNOW
Reasoning-response protocol	Give a brief natural response in the respondent's voice, followed by the coded answer. Uncertainty or an informal comment is allowed. Do not give a formal medical explanation. Q# response: ... Q# final answer: YES/NO/DON'T KNOW

Notes: Angle-bracketed text denotes a runtime field inserted separately for each respondent and stage. The system message is supplied on every model call. Baseline calls include the respondent's recorded DHS answers to Q1–Q3; post-exposure calls include the recorded answers to Q1–Q7 and, where applicable, the scenario-specific media feed.

Table 20: Variables Visible in the Respondent Profile

Profile variable	Meaning supplied to the model
`country_context`	Senegal, mid-to-late 2010s.
`gender`	Respondent's sex/gender as recorded in the sampling frame.
`age`	Respondent's age.
`age_group`	Respondent's age-group category, where available.
`region`	Administrative region or department name from the sampling frame.
`region_type`	Rural/urban classification.
`education`	Recoded educational-attainment category.
`education_years_raw`	Original education-years value from the sampling frame.

Continued on next page

Table 20 – continued from previous page

Profile variable	Meaning supplied to the model
`literacy`	Recorded literacy category.
`wealth`	Recorded household wealth category or index group.
`marital_status_raw`	Original marital-status value from the sampling frame.
`radio_access`	Recorded access to or ownership of a radio.
`tv_access`	Recorded access to or ownership of a television.
`baseline_internet_access`	Recoded prior internet-use category, such as no prior use or some previous use.
`antenatal_care_visits_raw`	Original antenatal-care visits value.
`health_worker_visit`	Recorded health-worker-visit measure.
`hiv_tested_during_antenatal_raw`	Original indicator for HIV testing during antenatal care.

Notes: The model does not receive the iteration number, agent ID, treatment/group assignment, treatment indicator, response mode as a profile variable, randomisation stratum, baseline scores, raw DHS answer fields, or other hidden experimental details. The recorded HIV answers are nevertheless shown separately in the baseline and post-exposure prompt blocks described above.

A.3 Sensitivity to the cure question

Question 6 asks, "At present, can HIV be cured?" and scores "no" as correct under the intended routine-care interpretation. Routine antiretroviral treatment suppresses HIV but does not eradicate it. Exceptional cure reports have followed stem-cell transplantation undertaken to treat life-threatening cancers. These procedures carry substantial risks and are not a routine treatment for HIV.

Some feed items discuss these exceptional cases to represent ambiguity in online health information. An accurate report of an exceptional cure can still invite an unsupported inference about the availability of routine treatment. However, the question does not explicitly distinguish these interpretations. A "yes" response may therefore reflect recognition of an exceptional case rather than a mistaken belief about routine care. The primary three-question DHS outcome excludes Q6. Table **??** confirms that excluding Q6 preserves the ordering of the information environments and the negative effect of switching from traditional media to the noisy online feed.

A.4 Stability across iterations

Table 21 estimates the pooled ANCOVA specification separately for each of the five 600-profile iterations. The ordering of Scenarios I–IV is preserved in every iteration for both outcome indices. Scenario IV remains positive for the DHS-3 outcome, with estimates from 0.060 to 0.097, but negative for the full-seven outcome, with estimates from -0.037 to -0.008. Scenario V is negative throughout: its coefficients range from -0.122 to -0.073 for DHS-3 and from -0.243 to -0.211 for the full-seven index. The pooled findings therefore do not depend on one profile draw or one iteration of model calls.

Table 21: Pooled ANCOVA Treatment Effects Across Iterations

Panel A: DHS-3 outcome						
Treatment arm	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5	Pooled
Scenario I: Traditional media	0.216	0.198	0.213	0.178	0.167	0.194
Scenario II: Online, informative	0.189	0.197	0.177	0.099	0.162	0.165
Scenario III: Online, mixed	0.138	0.133	0.118	0.098	0.105	0.118
Scenario IV: Online, noisy	0.097	0.092	0.069	0.060	0.071	0.078
Scenario V: Media substitution, I to IV	-0.103	-0.122	-0.122	-0.073	-0.114	-0.107
Panel B: Full-seven outcome						
Treatment arm	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5	Pooled
Scenario I: Traditional media	0.200	0.199	0.204	0.197	0.186	0.197
Scenario II: Online, informative	0.145	0.165	0.144	0.113	0.156	0.144
Scenario III: Online, mixed	0.076	0.069	0.056	0.060	0.064	0.065
Scenario IV: Online, noisy	-0.014	-0.008	-0.037	-0.036	-0.021	-0.023
Scenario V: Media substitution, I to IV	-0.225	-0.243	-0.227	-0.211	-0.236	-0.228

Notes: Each iteration draws a fresh pooled sample of 300 treated and 300 control respondents. Cells report pooled ANCOVA treatment coefficients. Within-iteration ANCOVA models condition on the pre-treatment outcome, literacy, education, rural or urban residence, and response mode; iteration fixed effects are omitted because they do not vary within an iteration. The final column pools all five iterations and additionally conditions on iteration. Scenarios I–IV compare the stated environment with no new information; Scenario V compares the I-to-IV media switch with continued traditional media.

A.5 Question-level effects

Table 22 decomposes the two indices into their seven component questions. Under Scenario IV, the pregnancy and delivery outcomes rise by 0.104 and 0.152 relative to no additional information, while the breastfeeding estimate is small and statistically insignificant. The same feed reduces correct responses to the mosquito and kissing questions by 0.259 and 0.324. Scenario V produces negative estimates for six of the seven questions, including declines of 0.392 for mosquitoes and 0.589 for kissing. These component estimates show that the full-seven decline is concentrated in questions directly contradicted by misleading items, whereas some maternal-transmission facts remain learnable from the noisy feed when the counterfactual is no new information.

Table 22: Pooled ANCOVA Question-Level Treatment Effects

Question	I	II	III	IV	V
Q1: Mother-to-child transmission during pregnancy	0.210***	0.198***	0.152***	0.104***	-0.095***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Q2: Mother-to-child transmission during delivery	0.206***	0.203***	0.175***	0.152***	-0.038***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)
Q3: Mother-to-child transmission during breastfeeding	0.167***	0.093***	0.028*	-0.022	-0.192***
	(0.014)	(0.015)	(0.015)	(0.014)	(0.016)
Q4: Healthy-looking person can have HIV/AIDS	0.183***	0.152***	0.176***	0.173***	0.011
	(0.015)	(0.015)	(0.015)	(0.014)	(0.016)
Q5: Mosquito bites do not spread HIV/AIDS	0.126***	0.044***	-0.086***	-0.259***	-0.392***
	(0.007)	(0.010)	(0.012)	(0.013)	(0.013)
Q6: There is no cure for HIV/AIDS	0.268***	0.210***	0.127***	0.020*	-0.290***
	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)
Q7: Kissing does not spread HIV/AIDS	0.226***	0.120***	-0.111***	-0.324***	-0.589***
	(0.008)	(0.011)	(0.014)	(0.015)	(0.013)
Observations	3,000	3,000	3,000	3,000	3,000

Notes: Cells report pooled ANCOVA treatment estimates; HC2 heteroskedasticity-robust standard errors are in parentheses. The pooled sample combines direct and brief-reasoning response protocols and conditions on response mode, baseline knowledge, iteration, literacy, education, and rural or urban residence. Scenarios I–IV compare the stated information environment with no new information. Scenario V compares the I-to-IV media switch with continued traditional media, conditioning on the post-traditional outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.6 Sampling-temperature sensitivity

Table 23 repeats the reasoning protocol at temperatures 0.3, 0.5, and 0.7 while retaining the same profiles, assignments, feeds, top-p, and context window. The ranking of Scenarios I–IV is unchanged at each temperature. Scenario IV remains positive for DHS-3 and negative for the full-seven index, while Scenario V remains negative for both outcomes. The Scenario V coefficient ranges from -0.117 to -0.129 for DHS-3 and from -0.233 to -0.244 for the full-seven index. The central contrasts are therefore stable over the reported change in sampling randomness within the same model checkpoint.

Table 23: ANCOVA Treatment Effects by Sampling Temperature

Outcome	I	II	III	IV	V
Panel A: Temperature 0.3					
DHS-3 outcome	0.202***	0.162***	0.124***	0.075***	-0.117***
	(0.018)	(0.017)	(0.017)	(0.016)	(0.017)
Full-seven outcome	0.182***	0.123***	0.047***	-0.047***	-0.236***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)
Panel B: Temperature 0.5					
DHS-3 outcome	0.238***	0.160***	0.130***	0.092***	-0.120***
	(0.016)	(0.017)	(0.016)	(0.015)	(0.017)
Full-seven outcome	0.203***	0.123***	0.052***	-0.043***	-0.244***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.010)
Panel C: Temperature 0.7					
DHS-3 outcome	0.238***	0.176***	0.124***	0.102***	-0.129***
	(0.016)	(0.017)	(0.017)	(0.016)	(0.015)
Full-seven outcome	0.207***	0.131***	0.050***	-0.035***	-0.233***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)

Notes: Cells report reasoning-protocol ANCOVA estimates. Standard errors are HC2 heteroskedasticity-robust and shown in parentheses. Scenarios I–IV compare the assigned information environment with no new information; Scenario V compares a traditional-to-online switch with continued traditional media. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Natural-Experiment Appendix

This appendix documents survey coverage and variable construction, then reports estimates that assess the composition of the sample, treatment assignment, inference, comparison geography, falsification outcomes, and heterogeneity.

B.1 Survey coverage, variables, and treatment construction

Table 24 lists the DHS fieldwork years used in each country and reports respondent counts separately for women and men. Coverage runs from 2003 to 2023, with more frequent observations for Senegal and fewer usable rounds in the other five countries. The sex-specific counts also clarify why women form the benchmark sample and men provide a supplementary comparison.

Table 24: DHS Survey Years by Country

Country x Year	Burkina Faso	Ghana	Guinea	Mali	Senegal	Sierra Leone
	2003	2003	2005	2006	2005	2008
	2010	2008	2012	2012	2010	2013
	2021	2014	2018	2018	2014	2019
		2022		2023	2015	
					2016	
					2017	
					2023	
Female N	34687	24219	24870	34815	74325	34628
Male N	10217	13578	10398	11845	16210	25297
Total N	44904	37797	35268	46660	90535	59925

Table 25 records the country-rounds that enter at least one Afrobarometer analysis. The available questions change across rounds, so the estimation samples differ across internet-use, news-use, and trust outcomes. The table identifies the fieldwork years underlying each part of that evidence and distinguishes country-round coverage from regression-specific observation counts.

Table 25: Afrobarometer Survey Coverage Across Reported Analyses

Country	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Round 9	Round 10	Total N
Burkina Faso			2008	2012	2015	2017	2019	2022		7,133
Ghana	2002	2005	2008	2012	2014	2017	2019	2022	2024	17,855
Guinea				2013	2015	2017	2019	2022	2024	7,140
Mali	2002	2005	2008	2013	2014	2017	2020	2022	2024	10,846
Senegal	2002	2005	2008	2013	2014	2017	2021	2022		9,557
									Total observations	52,531

Notes: Cells report fieldwork years for Afrobarometer country-rounds that contribute to at least one reported analysis. Country totals count all records in these contributing rounds; regression-specific observation counts are smaller because question and control availability varies across outcomes. Rounds 2 and 3 supply the pre-rollout media-trust outcomes, while Round 4 contributes to the traditional-news and institutional-trust analyses. The preferred internet-use specification uses Rounds 5–10 and contains 38,831 observations.

Table 26 gives the exact coding of the DHS, GIS, and Afrobarometer variables. It records the preferred binary fibre treatment, the road restriction, the inverse-distance measures, the maternal HIV outcomes, mechanism variables, controls, and fixed-effect identifiers. The table also states how invalid, refused, and "don't know" responses are handled.

Table 26: Variable Definitions

Variable name	Source	Definition
Panel A: Demographic and Health Surveys		
`cables_15_1`	DHS/GIS	Indicator equal to one when an enumeration area lies within 15 kilometres of a fibre-network node or within 1 kilometre of a fibre cable.
`cables_15_1_post`	DHS/GIS	Preferred treatment indicator: `cables_15_1` multiplied by an indicator for the survey year being at or after the local network-connection year. The timing convention is uniform across countries.
`roads_10km`	DHS/GIS	Indicator equal to one when an enumeration area lies within 10 kilometres of a main road. The main road-restricted sample retains treatment-zone observations or observations for which this indicator equals one.
`inv_dist_cable_post6`	DHS/GIS	Inverse distance to the nearest fibre cable, scaled between zero and one and interacted with the relevant post-connection indicator.

Continued on next page

Table 26 continued

Variable name	Source	Definition
inv_dist_ cablenode_post6	DHS/GIS	Inverse distance to the nearest fibre cable or node, scaled between zero and one and interacted with the relevant post-connection indicator.
aids_preg_raw	DHS	Knowledge that HIV can be transmitted from mother to child during pregnancy: 0=incorrect, 1="don't know", and 2=correct.
aids_del_raw	DHS	Knowledge that HIV can be transmitted from mother to child during delivery: 0=incorrect, 1="don't know", and 2=correct.
aids_brst_raw	DHS	Knowledge that HIV can be transmitted from mother to child through breastfeeding: 0=incorrect, 1="don't know", and 2=correct.
aids_maternal	DHS	Four-item maternal HIV/AIDS knowledge index, equal to the mean of indicators for knowledge of general mother-to-child transmission and transmission during pregnancy, delivery, and breastfeeding. The index ranges from zero to one.
dk_preg	DHS	Indicator equal to one when the respondent answers "don't know" to the pregnancy-transmission question.
dk_del	DHS	Indicator equal to one when the respondent answers "don't know" to the delivery-transmission question.
dk_brst	DHS	Indicator equal to one when the respondent answers "don't know" to the breastfeeding-transmission question.
u384a	DHS	Indicator that the respondent heard a family-planning message on the radio during the preceding few months.
u384b	DHS	Indicator that the respondent heard a family-planning message on television during the preceding few months.
u384c	DHS	Indicator that the respondent read a family-planning message in a newspaper or magazine during the preceding few months.
hw_visit	DHS	Indicator that a family-planning fieldworker visited the respondent during the preceding 12 months.
ant_visits	DHS	Number of antenatal-care visits reported for the relevant pregnancy; values of 20 or more are coded as 20.
ant_visits_cap10	DHS	Alternative antenatal-visit measure used for sensitivity analysis; values of 10 or more are coded as 10.
final_say_1	DHS	Who has the final say on making large household purchases
final_say_2	DHS	Who has the final say on making visits to family or relatives
domestic_violence_1	DHS	Wife beating is justified if she neglects the children
domestic_violence_2	DHS	Wife beating is justified if she argues with him (husband)
age	DHS	Respondent's age in completed years.
gender	DHS	Respondent's sex: 1=male and 2=female.
marriage	DHS	Current marital-status category: never in union, currently in union, or formerly in union.
hhsize	DHS	Number of de facto household members.
literacy	DHS	Ordered literacy category: cannot read, can read part of a sentence, or can read a complete sentence.
educ	DHS	Completed years of education.
radio	DHS	Indicator that the household owns a radio.
tv	DHS	Indicator that the household owns a television.
wealth	DHS	DHS household wealth quintile, ranging from 1 (poorest) to 5 (richest).

Continued on next page

Table 26 continued

Variable name	Source	Definition
rururb	DHS	Place of residence: 1=urban and 2=rural.
electric	DHS	Indicator that the household has electricity.
grid_id6	DHS/GIS	Identifier for a country-qualified 10-by-10-kilometre grid cell; used for the preferred spatial fixed effects and clustering.
region_year_id6	DHS/GIS	Identifier for a country-qualified ADM1 region by DHS survey year; used for the preferred region-by-year fixed effects.
Panel B: Afrobarometer		
treat1	Afro/GIS	Indicator equal to one when the interview occurs after the mapped local fibre-network connection year.
ever_treated	Afro/GIS	Indicator that the respondent's location is eventually reached by the mapped fibre network; used in pre-rollout comparisons.
internet_use	Afrobarometer	Frequency of internet use: 0=never, 1=less than once a month, 2=a few times a month, 3=a few times a week, and 4=every day.
radio_news	Afrobarometer	Frequency of obtaining news from radio, coded from 0 (never) to 4 (every day).
tv_news	Afrobarometer	Frequency of obtaining news from television, coded from 0 (never) to 4 (every day).
internet_news	Afrobarometer	Frequency of obtaining news from the internet, coded from 0 (never) to 4 (every day).
sm_news	Afrobarometer	Frequency of obtaining news from social media, coded from 0 (never) to 4 (every day).
trust_pres	Afrobarometer	Trust in the president, coded from 0 (not at all) to 3 (a lot).
trust_ruling	Afrobarometer	Trust in the ruling party, coded from 0 (not at all) to 3 (a lot).
trust_local	Afrobarometer	Trust in local government, coded from 0 (not at all) to 3 (a lot).
trust_parl	Afrobarometer	Trust in parliament, coded from 0 (not at all) to 3 (a lot).
trust_electoral	Afrobarometer	Trust in the electoral commission, coded from 0 (not at all) to 3 (a lot).
trust_pub	Afrobarometer	Ghana-specific trust in public or government media, coded from 0 (not at all) to 3 (a lot).
trust_priv	Afrobarometer	Ghana-specific trust in private media, coded from 0 (not at all) to 3 (a lot).
trust_sm	Afrobarometer	Trust in social media, coded from 0 (not at all) to 3 (a lot).
cor_pub	Afrobarometer	Perceived corruption in public or government media, with higher values indicating that corruption is perceived to be more widespread.
cor_priv	Afrobarometer	Perceived corruption in private media, with higher values indicating that corruption is perceived to be more widespread.
trust_gov_broad	Afrobarometer	Pre-rollout trust in government broadcasting, harmonised across Afrobarometer rounds 2 and 3.
trust_ind_broad	Afrobarometer	Pre-rollout trust in independent broadcasting, harmonised across Afrobarometer rounds 2 and 3.
trust_gov_paper	Afrobarometer	Pre-rollout trust in government newspapers, harmonised across Afrobarometer rounds 2 and 3.
trust_ind_paper	Afrobarometer	Pre-rollout trust in independent newspapers, harmonised across Afrobarometer rounds 2 and 3.
age	Afrobarometer	Respondent's age in completed years.
gender	Afrobarometer	Respondent's sex: 1=male and 2=female.

Continued on next page

Table 26 continued

Variable name	Source	Definition
`educ`	Afrobarometer	Highest education category, ranging from 0 (no formal schooling) to 9 (postgraduate education).
`wo_food`	Afrobarometer	Frequency during the preceding year that the respondent's household went without enough food, coded from 0 (never) to 4 (always).
`wo_income`	Afrobarometer	Frequency during the preceding year that the respondent's household went without a cash income, coded from 0 (never) to 4 (always).
`urbrur`	Afrobarometer	Place of residence: 1=urban and 2=rural.
`mobile`	Afrobarometer	Indicator that the respondent or household has access to a mobile phone.
`radio`	Afrobarometer	Indicator that the household owns a radio.
`tv`	Afrobarometer	Indicator that the household owns a television.
`country_region`	Afrobarometer	Country-qualified region identifier used for the preferred spatial fixed effects.
`country_year`	Afrobarometer	Country-by-survey-year identifier used for the preferred time fixed effects.
`country_town`	Afrobarometer	Country-qualified town or village identifier used for clustering standard errors.

Notes: Invalid, refused, and "don't know" codes are set to missing unless "don't know" is itself the outcome. Most notably, this is for the mother to child HIV transmission questions, where don't know is coded as 1. DHS fixed-effect identifiers are country-qualified. Afrobarometer variables harmonise the corresponding questions across survey rounds; valid categories retain their original ordering.

Table 27 repeats the internet-use specification for women. The fibre-proximity coefficient remains positive and similar to the pooled estimate. The remaining Afrobarometer estimates retain the mixed sample to preserve precision in the much smaller samples; those estimates describe changes in the broader information environment, rather than changes specific to women.

Table 27: Fibre Proximity and Internet Use: Women Only

	Average effect	Without Cash Income	Implied effect Subgroup total
Fibre proximity	0.174***		
	(0.0243)		
Fibre proximity: never (reference)		0.337***	0.337***
		(0.0461)	(0.0461)
Fibre proximity × once or twice		0.110	0.446***
		(0.0682)	(0.0597)
Fibre proximity × several times		-0.0768	0.260***
		(0.0557)	(0.0410)
Fibre proximity × many times		-0.307***	0.0300
		(0.0524)	(0.0328)
Fibre proximity × always		-0.350***	-0.0128
		(0.0552)	(0.0369)
Controls			
Age	-0.00925***	-0.00920***	
	(0.000596)	(0.000596)	
Female	Omitted	Omitted	
Education	0.246***	0.244***	
	(0.00555)	(0.00553)	
Without food	-0.0246***	-0.0261***	
	(0.00667)	(0.00665)	
Without cash income	-0.0596***		
	(0.00735)		
Rural	-0.248***	-0.247***	
	(0.0256)	(0.0254)	
Country-region FE	Yes	Yes	
Country-by-year FE	Yes	Yes	
Mean dependent variable	0.641	0.641	
Observations	19,354	19,354	
R^2	0.393	0.396	

Notes: The sample is restricted to women; gender is omitted because it is constant. Internet use, measured from zero (never) to four (every day). "Fibre proximity" is the treatment indicator derived from the mapped fibre network. Levels 0–4 refer to the frequency of going without cash income in the past 12 months. Column (1) reports the average-effect specification, with a linear without cash income control. Column (2) includes without cash income category indicators and their interactions with fibre proximity. Both specifications include region and country-by-year fixed effects. Standard errors, clustered by town/village, are in parentheses; subgroup totals use delta-method standard errors. Stars use the reported regression t-test p-values in columns (1)–(2) and the reported z-test p-values for subgroup totals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Pooled and male estimates

Table 28 combines women and men. The preferred pregnancy coefficient is −0.047 (SE 0.020), compared with −0.099 in the benchmark sample of women. The pooled estimates are also negative for delivery, breastfeeding, and the maternal index; the index falls by 3.2 percentage points (SE 0.011). The smaller pooled coefficients reflect the different pattern among men rather than instability as controls enter the pregnancy specification, where the estimate changes from −0.051 to −0.047.

Table 28: Internet Access and Maternal HIV/AIDS Knowledge: Full Sample

	Pregnancy				Delivery	Breastfeeding	Maternal index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment							
Fibre proximity	-0.051**	-0.049**	-0.048**	-0.047**	-0.059***	-0.037**	-0.032***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.022)	(0.019)	(0.011)
Demographic controls							
Age		0.002***	0.003***	0.003***	0.004***	0.001***	0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female		-0.029**	-0.018	-0.018	0.012	0.099***	0.030***
		(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.008)
Married		0.019***	0.024***	0.024***	0.032***	0.022***	0.026***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)
Household size		-0.001**	-0.001*	-0.001*	-0.001**	-0.000	-0.000*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Socioeconomic controls							
Literacy			0.012***	0.012***	0.007*	-0.006*	0.013***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Education			0.004***	0.004***	0.011***	0.001	0.006***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Radio			0.012**	0.011**	0.009**	-0.003	0.009***
			(0.005)	(0.005)	(0.004)	(0.004)	(0.002)
Television			-0.006	0.003	0.006	-0.002	0.005
			(0.006)	(0.007)	(0.006)	(0.007)	(0.003)
Wealth			0.000	0.002	0.008**	-0.001	0.004**
			(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Geographic controls							
Rural				0.005	0.005	0.014	0.001
				(0.011)	(0.011)	(0.012)	(0.006)
Electricity				-0.020***	-0.003	-0.018**	-0.004
				(0.008)	(0.007)	(0.007)	(0.004)
Constant	1.517***	1.478***	1.410***	1.400***	1.253***	1.229***	0.482***
	(0.006)	(0.027)	(0.030)	(0.034)	(0.037)	(0.039)	(0.020)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	No	No	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	No	Yes	Yes	Yes	Yes
Sample	Full sample	Full sample	Full sample	Full sample	Full sample	Full sample	Full sample
Mean dependent variable	1.502	1.500	1.500	1.500	1.489	1.443	0.661
Observations	237088	235070	232550	232517	232507	232510	237909
R^2	0.083	0.087	0.088	0.088	0.085	0.098	0.103

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The sample includes women and men from the six countries. The dependent variables and maternal index are defined as in the main table. All specifications use the six-country sample. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 29 reports positive static estimates for men. Fibre proximity raises the pregnancy score by 0.082 points (SE 0.030), the delivery score by 0.071 points (SE 0.031), and the breastfeeding score by 0.062 points (SE 0.029). The corresponding maternal-index estimate is 0.051 (SE 0.015). These regressions use the same road-restricted comparison and fixed-effect structure as the benchmark estimates for women.

Table 29: Internet Access and Maternal HIV/AIDS Knowledge: Men

	Pregnancy	Delivery	Breast-feeding	Maternal index
	(1)	(2)	(3)	(4)
Fibre proximity	0.082***	0.071**	0.062**	0.051***
	(0.030)	(0.031)	(0.029)	(0.015)
Grid fixed effects	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes
Sample	Men	Men	Men	Men
Mean dependent variable	1.535	1.510	1.395	0.659
Observations	68697	68695	68698	70010
R^2	0.157	0.168	0.154	0.187

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The dependent variables measure pregnancy, delivery, and breastfeeding transmission knowledge; the maternal index combines the general mother-to-child transmission item and the three route-specific indicators. The sample is men. All specifications use the six-country sample. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5 shows why the male coefficients receive a more cautious interpretation. The average effect over event years +1 through +5 is 0.072 (SE 0.033; $p = 0.027$) under conventional TWFE but 0.051 (SE 0.033; $p = 0.119$) in the stacked design. The stacked design weakly supports a positive post-connection estimate of 0.0467 ($p = 0.093$) for the whole sample duration. The stacked pre-treatment coefficients also fluctuate: two leads are individually significant, and their joint test rejects at the 10 per cent level ($p = 0.067$). The positive static coefficients therefore do not carry over to a precise post-treatment average under the preferred timing design.

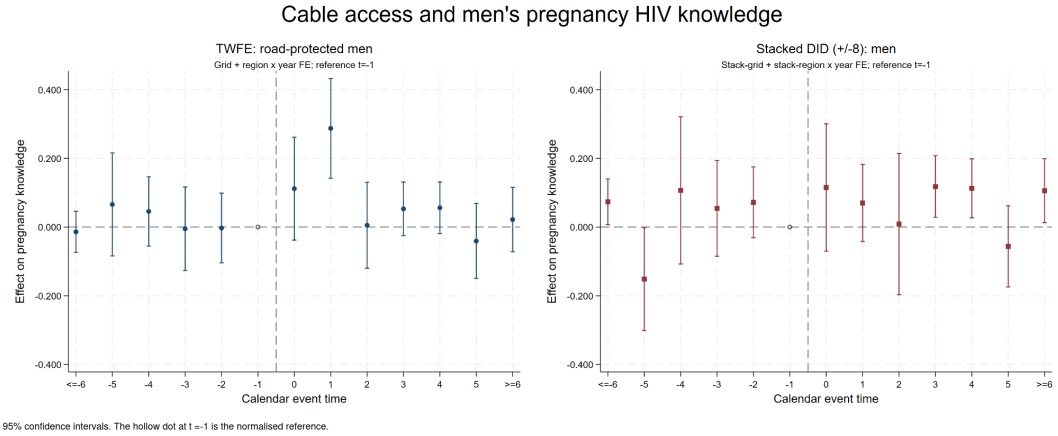


Figure 5: TWFE and stacked event-study estimates for men's pregnancy-transmission knowledge

Notes: Both panels use the road-restricted sample. The stacked specification applies cohort-specific eight-year windows and clean-control restrictions. Bars show 95 per cent confidence intervals; event time −1 is normalised to zero.

Table 30 further shows that the male estimates depend on the spatial treatment measure. The preferred and cable-only binary definitions produce positive pregnancy coefficients of 0.082 and 0.075, but the estimate falls to 0.026 and becomes statistically insignificant under the narrower 10-kilometre node rule. Inverse cable distance produces positive and statistically significant estimates for pregnancy knowledge and the maternal index. The cable-and-node measure is less stable: its pregnancy coefficient is statistically insignificant in both samples, and its index coefficient is significant only at the 10 per cent level in the road-restricted sample. This sensitivity, together with the event-study evidence, makes the male results an unreliable comparison, rather than a counterpart to the benchmark finding for women.

Table 30: Robustness to Alternative Treatment Definitions: Men

	Alternative treatment radii			Inverse-distance specifications			
	15 km node + 1 km cable	15 km cable only	10 km node + 1 km cable	Cable distance		Cable-node distance	
	Pregnancy	Pregnancy	Pregnancy	Pregnancy	Maternal index	Pregnancy	Maternal index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road-restricted sample							
Fibre proximity	0.082***	0.075**	0.026				
	(0.030)	(0.032)	(0.026)				
Inverse distance				0.117***	0.047**	0.057	0.035*
				(0.036)	(0.019)	(0.040)	(0.019)
Unrestricted sample							
Fibre proximity	0.081***	0.071**	0.029				
	(0.028)	(0.030)	(0.025)				
Inverse distance				0.115***	0.038**	0.054	0.021
				(0.036)	(0.019)	(0.038)	(0.018)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample consists of men. The road-restricted sample retains all observations in each treatment zone, including pre-treatment observations, and restricts non-treatment-zone comparisons to locations within 10 kilometres of a main road. The unrestricted sample does not condition inclusion on proximity to a main road. Columns (1)–(3) report alternative binary treatment definitions. Column (1) uses the preferred definition: within 15 kilometres of a network node or within 1 kilometre of a fibre cable. Column (2) uses a 15-kilometre cable-only treatment. Column (3) uses a 10-kilometre node radius and a 1-kilometre cable radius. Columns (4)–(7) replace the binary treatment with inverse-distance measures scaled between zero and one. Columns (4) and (5) use inverse distance to fibre cables; columns (6) and (7) use inverse distance to cables and the 15-kilometre node radius. The four-item maternal HIV/AIDS knowledge index combines the general mother-to-child transmission item with pregnancy, delivery, and breastfeeding transmission knowledge. All specifications include the main controls, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.3 Attenuation over time check

Table 31 examines whether the movement toward zero at longer horizons reflects changes in adoption-year composition. In the full Q-corrected sample, the early-to-late difference is significant for pregnancy knowledge ((p=0.032)), but not for the maternal index

Table 31: Early and Long-Horizon Fibre Effects with Fixed Adoption-Year Cohorts

Specification	Early effect	Long effect	Early − long	p_Δ	p_{pre}	Stacked N
Pregnancy-transmission knowledge						
Full, Q; +6+	−0.095 (0.030)	0.011 (0.048)	−0.105 (0.049)	0.032	0.039	379,535
Common, Q size; +6+	−0.129 (0.064)	0.066 (0.057)	−0.195 (0.067)	0.003	0.029	169,088
Common, Q size; +6 only	−0.184 (0.062)	0.075 (0.060)	−0.259 (0.082)	0.002	< 0.001	169,088
Common, Q equal; +6+	−0.181 (0.102)	0.137 (0.081)	−0.318 (0.100)	0.002	0.029	169,088
Common, arm size; +6+	−0.105 (0.053)	0.034 (0.057)	−0.139 (0.057)	0.014	0.078	169,088
Maternal HIV knowledge index						
Full, Q; +6+	−0.056 (0.016)	−0.018 (0.024)	−0.038 (0.025)	0.125	0.003	384,645
Common, Q size; +6+	−0.059 (0.024)	0.020 (0.026)	−0.078 (0.026)	0.003	0.005	171,570
Common, Q size; +6 only	−0.111 (0.021)	0.028 (0.026)	−0.138 (0.031)	< 0.001	< 0.001	171,570
Common, Q equal; +6+	−0.068 (0.035)	0.062 (0.033)	−0.130 (0.034)	< 0.001	0.005	171,570
Common, arm size; +6+	−0.053 (0.022)	−0.003 (0.025)	−0.050 (0.025)	0.045	0.039	171,570

Notes: The road-restricted stacked design uses clean controls, an eight-year window, stack-by-grid and stack-by-region-by-year fixed effects, the paper's controls, and grid-clustered standard errors (in parentheses). The full-sample early effect averages the five event-time coefficients at $+1, \ldots, +5$. The common-cohort rows hold adoption years 2011 and 2012 fixed. For 2011, the early effect averages observed years $+3, +4, +5$; for 2012, it averages $+1, +2$. There is no cohort observed at every year from +1 through +6. Size aggregation fixes the cohort weights at 0.713 (2011) and 0.287 (2012), based on the number of treated respondents in each cohort's latest observed pre-treatment survey; equal aggregation gives each cohort weight 0.5. Q weights are recalculated within the common-cohort stack; arm weights balance total treated and control mass within each stack. The 2011 cohort uses event time −1 as its reference, but 2012 has no treated observations at −1 and uses −2. The +6+ bin contains treated observations at +6 and +7 only. Each terminal-bin choice is separately estimated, so its early coefficient may differ. p_Δ tests equality of the early and long effects using the joint clustered covariance matrix; p_{pre} is the joint test of the estimable pre-treatment coefficients. Stacked N counts repeated appearances across stacks. The common-cohort comparison does not hold treated countries or grid cells fixed across horizons.

((p=0.125)). Holding the 2011 and 2012 adoption-year cohorts and their aggregation weights fixed, the early-minus-late contrasts are (-0.195) for pregnancy knowledge and (-0.078) for the maternal index (both (p=0.003)). The contrast therefore survives fixing adoption-year shares. It does not hold locations fixed: the 2012 cohort's early and (+6) observations come from different countries, with no treated grid cells in common. The common-cohort Q-corrected pre-treatment tests also reject for both outcomes ((p=0.029) and (p=0.005), respectively).

B.4 Treatment assignment and comparison geography

Table 32 replaces the binary treatment boundary with continuous proximity measures and restricts the sample to women living within 50 kilometres of the relevant fibre infrastructure. Inverse cable distance produces estimates of −0.078 for pregnancy knowledge and −0.045 for the maternal index, both significant at the 10 per cent level. Incorporating proximity to network nodes increases the magnitudes to −0.126 and −0.074, both significant at the 1 per cent level. The similarity to the corresponding estimates in the broader sample reduces concern that either the binary service radius or very remote comparison locations account for the main result.

Table 32: Inverse-Distance Treatments within 50 Kilometres

	Cable distance		Cable and node distance	
	Pregnancy	Maternal index	Pregnancy	Maternal index
	(1)	(2)	(3)	(4)
Treatment				
Inverse cable distance	-0.078*	-0.045*		
	(0.045)	(0.024)		
Inverse cable-and-node distance			-0.126***	-0.074***
			(0.035)	(0.017)
Grid fixed effects	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes
Sample	Women within 50 km	Women within 50 km	Women within 50 km	Women within 50 km
Mean dependent variable	1.486	0.662	1.487	0.662
Observations	159401	163282	161854	165816
R^2	0.123	0.134	0.124	0.136

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The sample is women within 50 kilometres of the relevant fibre infrastructure. Inverse distance is scaled between zero and one. All specifications use the six-country sample. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.5 Device-Ownership Controls in Afrobarometer News Regressions

Table 33: Fibre Proximity and News Consumption: Omitting Device Ownership Controls

	Radio	Television	Internet	Social media
	(1)	(2)	(3)	(4)
Fibre proximity	0.0373*	0.2304***	0.1200***	0.1307***
	(0.0221)	(0.0319)	(0.0198)	(0.0245)
Country-region FE	Yes	Yes	Yes	Yes
Country-by-year FE	Yes	Yes	Yes	Yes
Mean dependent variable	2.9954	2.0827	0.8321	1.0331
Observations	42,282	42,200	33,927	27,251
R-squared	0.1020	0.3617	0.4076	0.4126

Notes: Outcomes measure news use on a zero-to-four frequency scale (never to every day). All columns control for age, gender, education, food deprivation, income deprivation, and rural/urban residence, and include country-qualified region and country-by-survey-year fixed effects. Device ownership controls are omitted. Each column retains respondents with nonmissing ownership information for the corresponding device: radio in column (1), television in column (2), and mobile phone in columns (3)–(4). Standard errors, clustered by country-qualified town/village, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.6 Fixed effects, clustering, and national shocks

Table 34 varies the spatial fixed effects, time fixed effects, and clustering level. All seven pregnancy coefficients are negative. They range from −0.044 with ADM2 fixed effects to −0.101 with hexagonal fixed effects; six are significant at the 1 per cent level, and the specification with common survey-year fixed effects is significant at the 5 per cent level. The estimate is −0.092 with country-by-year fixed effects, close to the benchmark estimate of −0.099 with region-by-year fixed effects.

Table 34: Robustness to Fixed Effects and Clustering

	Main	Grid FE; hex cluster	Hex FE; hex cluster	Grid FE; ADM2 cluster	ADM2 FE; grid cluster	Country × year FE; grid cluster	Year FE; grid cluster
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fibre proximity	-0.099***	-0.099***	-0.101***	-0.099***	-0.044***	-0.092***	-0.064**
	(0.029)	(0.027)	(0.028)	(0.032)	(0.016)	(0.030)	(0.030)
Spatial fixed effects	Grid cell	Grid cell	Hexagon	Grid cell	ADM2	Grid cell	Grid cell
Time fixed effects	Region x year	Region x year	Region x year	Region x year	Region x year	Country x year	Year
Clustering	Grid cell	Hexagon	Hexagon	ADM2	Grid cell	Grid cell	Grid cell
Mean dependent variable	1.486	1.486	1.486	1.486	1.486	1.486	1.486
Observations	163785	163785	163785	163785	163786	163785	163785
R^2	0.125	0.125	0.127	0.125	0.089	0.114	0.114

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The outcome is pregnancy-transmission knowledge. The table reports the spatial fixed effects, time fixed effects, and clustering level used in each specification. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.7 Changes in subgroup characteristics

Table 35 examines whether fibre exposure is associated with changes in the characteristics used in the heterogeneity analysis. The regressions use the road-restricted pregnancy-knowledge sample, with grid and region-by-year fixed effects. Each specification retains the remaining baseline controls but excludes the characteristic being examined. The television-ownership and mean-literacy coefficients are small and statistically insignificant. Education increases by approximately 0.11 years, while the coefficient on the average wealth-quintile code is statistically insignificant. The category-specific estimates nevertheless reveal changes in the wealth distribution: the fourth-quintile share rises by 4.3 percentage points ($p = 0.003$), while the richest-quintile share falls by 3.3 points ($p = 0.012$). The poorest-quintile share falls by 2.1 points, significant only at the 10 per cent level. These conditional comparisons indicate that wealth-group composition can change around fibre connection. The heterogeneity estimates consequently describe differences across groups observed at interview.

Table 35: Fibre Exposure and Subgroup-Membership Measures

	Wealth						Other characteristics		
	Index	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Literacy	Education	Television
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fibre proximity	0.033	-0.021*	-0.014	0.025	0.043***	-0.033**	-0.008	0.106**	0.014
	(0.031)	(0.012)	(0.012)	(0.015)	(0.014)	(0.013)	(0.010)	(0.052)	(0.010)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women	Women	Women	Women	Women
Mean dependent variable	3.263	0.149	0.171	0.202	0.226	0.252	0.684	3.635	0.427
Observations	163785	163785	163785	163785	163785	163785	163785	163785	163785
R^2	0.762	0.416	0.259	0.185	0.196	0.582	0.725	0.747	0.661

Notes: Each column regresses the indicated characteristic on fibre proximity. The sample is road-restricted women with nonmissing pregnancy-transmission knowledge; outcome-specific missing values and controls may change the number of observations across columns. Wealth-quintile outcomes equal one for membership in the stated quintile. All regressions include grid and region-by-year fixed effects; the other covariates follow the corresponding specifications in the replication code. Standard errors clustered by grid are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.8 Country, city, and distance restrictions

Table 36 removes one country at a time from the road-restricted sample. The pregnancy estimates remain negative and statistically significant in all six regressions, ranging from -0.056 when Sierra Leone is excluded to -0.128 when Senegal is excluded. The maternal-index estimates range from -0.053 to -0.078. No single national rollout accounts for the pooled result, although the change in magnitude across exclusions shows that countries contribute unequally to the estimate.

Table 36: Country-by-Country Exclusion Robustness Test

	Pregnancy knowledge (1)	Maternal index (2)
Exclude Burkina Faso	−0.099***	−0.067***
	(0.029)	(0.014)
Exclude Ghana	−0.097***	−0.067***
	(0.031)	(0.015)
Exclude Guinea	−0.106***	−0.068***
	(0.032)	(0.015)
Exclude Mali	−0.111***	−0.071***
	(0.030)	(0.014)
Exclude Sierra Leone	−0.056**	−0.053***
	(0.024)	(0.014)
Exclude Senegal	−0.128***	−0.078***
	(0.043)	(0.021)

Notes: Each column reports the fibre-proximity treatment coefficient from a country-exclusion specification. The applicable country is omitted from each row. Standard errors, clustered at the grid-cell level, are in parentheses. All specifications include grid fixed effects and region × year fixed effects. It uses the road-restriction sample. The sample size is between 104-146,000, mean dependent variable is 1.46-1.54 and 0.65-0.69 for pregnancy knowledge and maternal index, respectively. The R^2 is between 0.118-0.151. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 37 keeps the road restriction fixed while sequentially excluding the capital, two largest cities, and three largest cities in each country. The pregnancy coefficient remains between −0.077 and −0.083 after these exclusions, compared with −0.099 in the full road-restricted sample. The estimates remain significant at the 1 per cent level.

Table 37: Fibre Proximity and Maternal HIV/AIDS Knowledge: Road-Protected City Exclusions

	Road baseline (1)	Excluding capital (2)	Excluding two largest (3)	Excluding three largest (4)
Fibre proximity	-0.099***	-0.081***	-0.083***	-0.077***
	(0.029)	(0.025)	(0.026)	(0.026)
Grid fixed effects	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes
Sample	Road-protected women	Road-protected women	Road-protected women	Road-protected women
Mean dependent variable	1.486	1.482	1.486	1.488
Observations	163785	146557	136521	128548
R^2	0.125	0.134	0.136	0.135

Notes: Every column uses the road-restricted sample, which retains all observations in the preferred fibre-proximity treatment zone, including pre-treatment observations, and restricts non-treatment-zone comparisons to locations within 10 kilometres of a main road. The outcome is pregnancy-transmission knowledge. All specifications use the six-country sample. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.9 Placebos and interview length

Table 38 estimates the benchmark specification using outcomes that proxy for respondent composition, local infrastructure, household decision-making, and attitudes toward domestic violence. The coefficients for age, marital status, water-collection time, two decision-making measures, and two domestic-violence attitudes are all statistically insignificant. Water-collection time speaks to broad local development and infrastructure programmes; it does not test for health-specific interventions.

Table 38: Falsification Tests: Composition, Local Infrastructure, and Attitudes

	Demographics		Infrastructure	Attitudes and decision-making			
	Age	Married	Water time	Final say #1	Final say #2	Domestic violence#1	Domestic violence#2
Fibre proximity	0.066	-0.005	-0.914	0.004	0.021	0.004	0.009
	(0.122)	(0.013)	(0.940)	(0.023)	(0.021)	(0.026)	(0.027)
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women	Women	Women
Mean dependent variable	27.374	1.261	14.457	0.564	0.414	0.541	0.532
Observations	188431	188431	175180	132184	131678	188389	188410
R^2	0.195	0.260	0.268	0.205	0.185	0.094	0.098

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The final-say outcomes are coded 0/1/2, as fully the husband's descision, a joint decision to only the wife's decision. The two questions are related to household purchases and visits to friends and relatives, respectively. The domestic-violence outcomes indicate whether wife beating is considered justified if they neglect the children, and if they argue back to their husband, respectively. These questions are scored on the 1-5 scale, from strongly disagree to strongly agree. The female questionnaire is done in private, away from any spouse. All specifications use the six-country sample. Grid fixed effects and region × year fixed effects are included as shown. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 39 examines whether survey duration accounts for the decline in pregnancy knowledge or the increase in "don't know" responses. In the subsample with non-missing interview duration, the pregnancy coefficient is -0.107 with no duration control and remains -0.107 after adding linear or quadratic duration terms. The uncertainty coefficient changes only from 0.026 to 0.027 and 0.025. Survey duration therefore does not account for either result in this subsample.

Table 39: Interview Length and Survey-Fatigue Robustness

	Pregnancy knowledge			Uncertainty		
	(1)	(2)	(3)	(4)	(5)	(6)
Fibre proximity	-0.107***	-0.107***	-0.107***	0.026***	0.027***	0.025**
	(0.030)	(0.030)	(0.030)	(0.010)	(0.010)	(0.010)
Interview length		0.00005	0.00041		-0.00052***	-0.00223***
		(0.00020)	(0.00062)		(0.00007)	(0.00025)
Interview length squared			-0.000003			0.000016***
			(0.000005)			(0.000002)
Interview-length controls	None	Linear	Quadratic	None	Linear	Quadratic
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Grid fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Women	Women	Women	Women	Women	Women
Mean dependent variable	1.488	1.488	1.488	0.110	0.110	0.110
Observations	133,964	133,964	133,964	133,964	133,964	133,964
R^2	0.134	0.134	0.134	0.132	0.133	0.134

Notes: Columns (1)–(3) use knowledge that HIV/AIDS can be transmitted from mother to child during pregnancy as the dependent variable. Columns (4)–(6) use an indicator for answering "don't know" to the pregnancy-transmission question. All columns use the road-restricted six-country sample and restrict estimation to observations with non-missing interview length. Interview length is measured in minutes and restricted to values at or below 95. Columns (2) and (5) include a linear interview-length term; columns (3) and (6) also include its square. All specifications contain the controls used in the main specification, grid fixed effects, and region × year fixed effects. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.10 Additional heterogeneity

Table 40 reports implied treatment effects by antenatal-care use. The estimates are small and statistically insignificant for women reporting no visits or one to three visits. They become progressively more negative among women reporting four to six visits (−0.093), seven to nine visits (−0.137), and at least ten visits (−0.212); all three estimates are significant at the 1 per cent level. Antenatal-care use may itself respond to fibre exposure, so these estimates describe heterogeneity by realised care use rather than effects for predetermined subgroups.

B.11 Stacked DiD robustness checks

Table 41 compares the TWFE estimates with clean-control stacked DiD, which excludes already-treated observations from the control group. Matched-support TWFE uses the unique original observations included in the stacked regressions, helping distinguish sample restrictions from changes in the estimator. The decline in health-worker contact and increase in pregnancy uncertainty remain statistically significant under stacking. The literacy and wealth comparisons also remain broadly similar, although the estimated knowledge effects are generally smaller. The radio and newspaper coefficients are smaller and statistically insignificant. For newspapers, the estimate changes from −0.010 to −0.003 under matched-support TWFE and is −0.004 under stacking, indicating that the difference is largely associated with sample support.

Table 40: Heterogeneous Effects by Antenatal Care Use

	Pregnancy knowledge
	(1)
Fibre proximity × no antenatal visits	0.013
	(0.050)
Fibre proximity × 1–3 antenatal visits	-0.024
	(0.031)
Fibre proximity × 4–6 antenatal visits	-0.093***
	(0.031)
Fibre proximity × 7–9 antenatal visits	-0.137***
	(0.038)
Fibre proximity × 10+ antenatal visits	-0.212***
	(0.047)
Grid fixed effects	Yes
Region × year fixed effects	Yes
Controls	Yes
Sample	Mothers
Mean dependent variable	1.507
Observations	79,049
R^2	0.143

Notes: The road-restricted sample retains every observation in the applicable fibre-proximity treatment zone, including its pre-treatment observations, and restricts non-treatment-zone comparison observations to locations within 10 kilometres of a main road. The dependent variable is the 0–2 pregnancy-transmission knowledge score. Each coefficient is the implied effect of fibre proximity for the stated antenatal-care subgroup, calculated after estimating the fully interacted model. Antenatal visits are grouped using the reported number of visits for the relevant pregnancy. All specifications include the main controls, 10-kilometre by 10-kilometre grid fixed effects, and region × year fixed effects. Because antenatal-care use may itself respond to fibre proximity, the estimates should be interpreted as descriptive heterogeneity rather than effects for predetermined subgroups. Standard errors, clustered at the grid-cell level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 41: Clean-Control Stacked DiD: Information Sources and Heterogeneity

	Original TWFE	Matched-support TWFE	Stacked DiD
Panel A. Information sources, contact, and uncertainty			
Family-planning information: radio	-0.028**	-0.019	-0.006
	(0.013)	(0.014)	(0.015)
Family-planning information: newspaper/magazine	-0.010*	-0.003	-0.004
	(0.005)	(0.005)	(0.004)
Health-worker contact, HIV-questionnaire sample	-0.076***	-0.075***	-0.048***
	(0.024)	(0.024)	(0.015)
Uncertain about pregnancy transmission	0.030***	0.030***	0.024***
	(0.009)	(0.009)	(0.009)
Panel B. Pregnancy knowledge by literacy			
No literacy	-0.075**	-0.084**	-0.053**
	(0.029)	(0.033)	(0.027)
Partial literacy	-0.141***	-0.148***	-0.116***
	(0.034)	(0.037)	(0.030)
Full literacy	-0.118***	-0.131***	-0.089***
	(0.029)	(0.033)	(0.029)
Equal effects across literacy: Wald test	$p < 0.001$	$p < 0.001$	$p < 0.001$
Partial versus no literacy	$p < 0.001$	$p < 0.001$	$p < 0.001$
Full versus no literacy	$p = 0.004$	$p = 0.002$	$p = 0.021$
Partial versus full literacy	$p = 0.241$	$p = 0.400$	$p = 0.195$
Panel C. Pregnancy knowledge by wealth quintile			
Wealth quintile 1	-0.006	-0.013	0.002
	(0.034)	(0.036)	(0.035)
Wealth quintile 2	-0.021	-0.028	-0.013
	(0.030)	(0.032)	(0.031)
Wealth quintile 3	-0.088***	-0.094***	-0.066**
	(0.030)	(0.033)	(0.029)
Wealth quintile 4	-0.133***	-0.145***	-0.105***
	(0.034)	(0.037)	(0.030)
Wealth quintile 5	-0.146***	-0.166***	-0.130***
	(0.033)	(0.039)	(0.035)
Equal effects across wealth: Wald test	$p < 0.001$	$p < 0.001$	$p = 0.004$

Notes: Coefficients are treatment effects; standard errors clustered by the original country-qualified grid are in parentheses. Panels B and C show implied effects within literacy and wealth groups. The road-restricted sample retains all fibre-zone observations and untreated observations within 10 km of a main road. Pregnancy knowledge is coded 0/1/2; Panel A outcomes are binary. Each model uses its own complete-case sample. Original and matched TWFE absorb grid and region-by-year effects. Stacked DiD uses clean controls within cohort-specific -8 to $+8$ windows, stack-grid and stack-region-by-year effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The baseline stacked specification assigns treated observations unit weight and weights control observations in stack g by N_g^T/N_g^C. This gives the treated and control groups equal total weight within each stack, but does not equalise the contribution of treatment cohorts or event times to the pooled coefficient. As a robustness check, we re-estimate the specification using the repeated-cross-section Q-weights proposed by Wing et al. (2024). These weights align the cohort and event-time composition of the treated and control samples. The resulting estimates are largely unchanged at the reported precision.

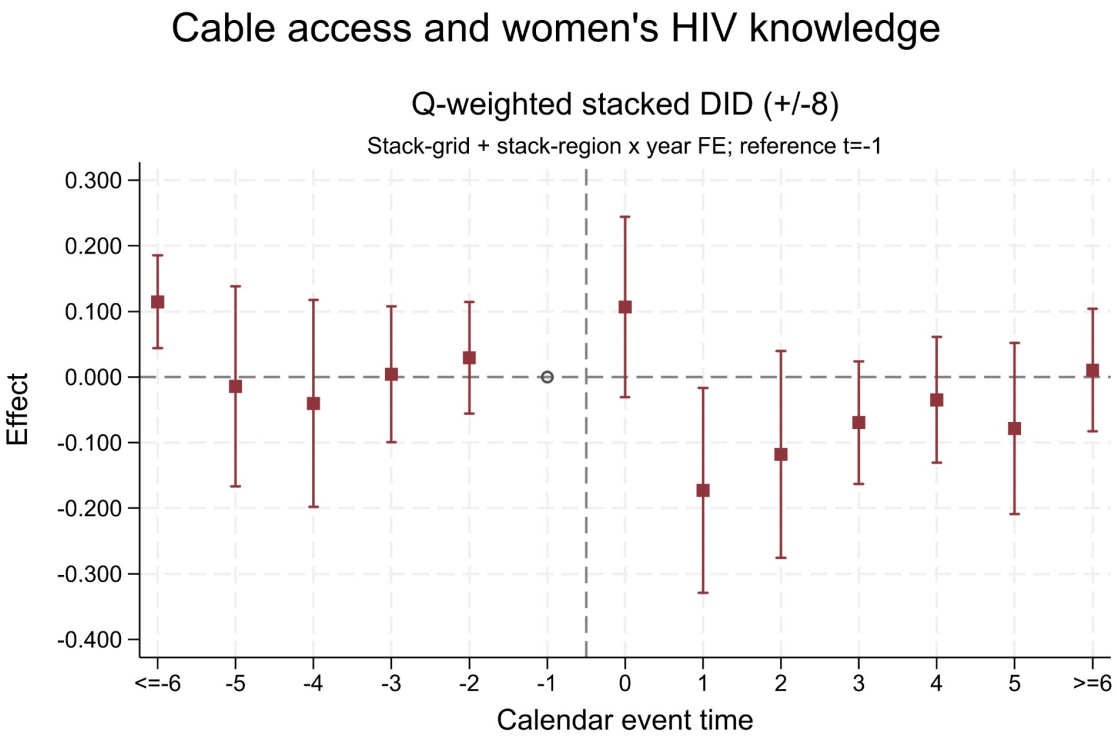


Figure 6: Q-weighted stacked event-study estimates for women's pregnancy-transmission knowledge