Published in final edited form as:

Am Sociol Rev. 2019 August 1; 84(4): 634–663. doi:10.1177/0003122419859574.

Data Collection as Disruption: Insights from a Longitudinal Study of Young Adulthood

Jeong Hyun Oh^a, Sara Yeatman^b, Jenny Trinitapoli^a

^aUniversity of Chicago

^bUniversity of Colorado Denver

Abstract

Research disrupts the social world, often by making respondents aware that they are being observed or by instigating reflection upon particular aspects of life via the very act of asking questions. Building on insights from the first Hawthorne studies, reflexive ethnographers, and methodologists concerned with panel conditioning, we draw on six years of research within a community in southern Malawi to introduce a conceptual framework for theorizing disruption in observational research. We present a series of poignant-yet-typical tales from the field and two additional tools—the refresher-sample-as-comparison and study-focused ethnography—for measuring disruption empirically in a longitudinal study. We find evidence of study effects in many domains of life that relate directly to our scope of inquiry (i.e., union formation, fertility) and in some that extend beyond it (i.e., health). Moreover, some study effects were already known and discussed in the broader community, which was also affected by our research in unintended ways. We conclude that the assumption of non-interactivity in observational research is shaky at best, urging data-gatherers and users to think more seriously about the role of disruption in their work.

Keywords

data collection; disruption; panel conditioning; longitudinal studies; behavioral adaptation; future-oriented thinking

This article addresses a long-standing question in social science research: How does research influence respondents? From the first Hawthorne studies to more recent concerns with panel conditioning in longitudinal surveys, one thing is clear: social-scientific inquiry is often disruptive. In experimental research, interventions are designed to do precisely that—alter a given course of events. But ostensibly observational studies also affect social life, particularly when respondents are aware of the behavior under observation. If the mere act of observing individuals can influence the behaviors researchers are trying to understand, might standard research practices not also alter other aspects of respondents' lives? Life beyond the purview of the study itself?

In this article, we ask *whether* and *how* research influences respondents, and we consider what this means for the data-gathering efforts prevalent in the social sciences, especially sociology. We approach these questions in three ways. First, drawing from multiple methodological traditions, we introduce the concept of disruption as a framework under which myriad concerns about the influences of research can be systematized and elaborated. Second, reflecting on six years of sustained engagement with a community in southern Malawi as part of the research process, we draw on fieldnotes documenting some of the ways our study disrupted respondents' lives, and we argue for elevating levels of transparency around disruptive forces in research. Third, we describe two systematically executed adjustments to our original fieldwork plan, designed specifically to measure the consequences of study participation on respondents and the community. This *refresher-sample-as-comparison* (RSC) method allows researchers to identify research effects in longitudinal studies, and study-focused ethnography allows us to see and hear how our research was discussed outside the walls of our research center.

Applied to the case of young adulthood in southern Malawi—a context characterized by extreme poverty, endemic HIV, early marriage, high fertility, and high levels of relationship instability—we find evidence that study participation influenced our respondents' thoughts and related behaviors. By asking questions about the future, we required our respondents to imagine those futures in specific ways—ways that may not get prompted very often in their everyday lives. In our study population, research-related questions not only increased future-oriented thinking generally, but also engendered a particular way of thinking about relationships, sex, and childbearing—the very topics we set out to understand. As people who care about preventing HIV, promoting health, reducing unintended fertility, and alleviating poverty, these disruptive, unintended consequences of research are overwhelmingly positive. But to sociologists who divvy up the explanatory world into real effects and artifacts, these findings will be troubling, representing yet another type of error for analysts to navigate. We close with suggestions for how sociologists might harness disruption to promote good social science.¹

BACKGROUND

Disruption and Interactivity: Insights from Ethnography

Ethnographers often take on active roles within their communities of interest—sometimes as a strategy for advancing their research and sometimes to meet obligations of reciprocity that enhance their own sense of belonging (Adler and Adler 1987). Srinivas (1976) noted that his presence as a Western-educated, modern, Indian gentleman planted the expectation among villagers that he would be a reliable source of guidance and advice. His assigned role as a counselor and pseudo-doctor had consequences for his access, enabling him, for example, to interact with young, unwed women against local taboo. It also spilled over to some

¹Our interest in disruption precedes Martin's (2016) recent essay calling for a more "nightmare-resistant sociology," which pinpoints, among other heroic assumptions sociologists regularly make, the fantasy of non-interactivity in data collection. Martin's insights and arguments are very much in-line with our core concerns and sharpened our thinking about how disruption matters beyond our particular case, as did Leahey's (2008) reminder that research is a human activity that inevitably involves social influences and interactivity. Martin (2016:538) remarked that "those who are most likely to recognize that we may in fact have an interactive role in the production of data seem least interested in doing anything about it." We hope to be filed with the notable exceptions.

community members; his cook, for example, landed a lucrative job after leaving the village. Working in the surgical wing of a U.S. hospital, Bosk (2003) famously documented his desire to be *useful*, first to legitimize his status and subsequently to contribute to the community he came to love. As he assumed these helping roles, serving as emissary of the outside world, outlet for dissent, and resident historian of the rotating house-staff, Bosk's sense of indebtedness led him to intervene more and more in the community, sometimes contravening the basic operating norms of the house-staffs. Stack (2008), likewise, reflected upon her role as occasional driver in a disadvantaged community, noting that her transportation services probably increased movement in and out of the community and enhanced the flow of information among her respondents. Her role as errand runner also had economic consequences, saving her respondents considerable sums of money they had been spending on cab fare.

The admirable levels of transparency scholars like Bosk, Srivinas, Stack, and others bring to their disruptive presence in the field is enlightening. Although acknowledgment of interactivity occurs frequently in ethnography, it has not yet become standard practice in other methodological traditions within the social sciences. As evidenced by these ethnographic accounts, researchers' roles are not, by default, unobtrusive. Researcher-respondent interactions invade respondents' daily lives and alter their actions, reactions, interactions, and performances in their everyday settings (Schwartz and Jacobs 1979). These transparent accounts of researcher-respondent interactions illustrate how disruption gets generated in ordinary research settings—occasionally intentionally² but most often accidentally. If we know from ethnography that single researchers influence the thoughts and behaviors of their respondents and interlocutors, how might quantitative researchers begin to think about the disruptive effects of multi-person research teams and large-scale studies?

Survey Research on Study Effects

Quantitative scholars conceptualize study effects using a vocabulary of error and bias. Item non-response may rise, for example, when respondents learn the skip patterns and figure out how to minimize survey burden. Likewise, attrition-related worries are exacerbated when busy and reluctant respondents drop out of a study over time, introducing selectivity biases that may or may not be correctable with post-stratification weights. Sensitive topics are thought to be susceptible to *social desirability bias*, but the precise nature of the relationship between research experience and social desirability bias remains unclear. On the one hand, reports may become more “honest” over time, as interviewers and respondents build rapport (Torche et al. 2012). On the other hand, reports may become less trustworthy over time, as respondents perceive and then conform to researchers' presumed expectations (Halpern-Manners, Warren, and Torche 2014).³

²Some branches of ethnography actively use disruption to solicit reactions from people. However, we classify this as an active manipulation of the scene that is more closely linked to interventions, rather than participant observation or survey research that do not intentionally seek to disrupt people's lives.

³In a study of marijuana use, petty thievery, and driving-under-the-influence in the United States, Halpern-Manners and colleagues (2014) found that inexperienced respondents were most likely to give socially desirable answers. Findings from several other high-quality studies point in a similar direction (Kreuter, Presser, and Tourangeau 2008; Moorman and Podsakoff 1992; Paulhus 1984).

Survey researchers have generated a lot of specialized knowledge about how to identify and address biases within surveys, but they tend to spend far less time thinking about how survey-based research may genuinely influence respondents' lives beyond the responses they give in the data. Still, we know study participation can lead to changes in respondents' attitudes, primarily by altering cognitive attention and motivation (Torche et al. 2012; Warren and Halpern-Manners 2012). Repeated contemplation of interview topics may crystallize attitudes on issues about which respondents were previously ambivalent. Evidence of this can be read from the fact that frequent participation in a survey tends to reduce within-person dispersion across outcomes over time (Kroh, Winter, and Schupp 2016; Wooden and Li 2014) and the incidence of "Don't Know" answers tends to decline with survey experience (Binswanger, Schunk, and Toepoel 2013; Struminskaya 2016). In other words, survey researchers know that respondents' attitudes and opinions get activated by the research process, just as they sometimes get activated by conversations and interactions that occur "naturally" in their lives (Christakis and Fowler 2013; Rutenberg and Watkins 1997).

Extending beyond attitudes, other scholars have suggested that the mere act of asking can change behavior by stimulating respondents' orientation to topics covered in the survey. Most of this evidence comes from the health-related social sciences and finds that both healthy and unhealthy behaviors can emerge from study participation, primarily by catalyzing an intention. Repeatedly asking questions about a certain behavior might serve as a reminder to act in a particular way (Williams, Block, and Fitzsimons 2006). Another channel via which research can lead to behavior change is through the provision of new information. In an intensive, longitudinal study of reproductive health in the United States, for example, young female students who were asked survey questions weekly came to adopt more effective contraceptive methods (Barber et al. 2016). In all likelihood, the study, which was in part designed to study contraceptive use, familiarized them with a wider range of contraceptive options than they had previously realized were available.

Whether resulting in actual behavioral changes or changes in reporting only, survey reactivity poses many questions, only a few of which have been answered conclusively. We know that when the inter-survey period is long, factors such as memory and life events override the effects of the prior survey on subsequent reports (Halpern-Manners, Warren, and Torche 2017; Levav and Fitzsimons 2006). We also know that survey modes and interviewer-respondent interactions can be designed to minimize survey reactivity and promote the collection of consistent and accurate answers to intrusive questions (Tourangeau and Smith 1996). But that is about it. Generally speaking, survey reactivity and panel conditioning remain poorly understood phenomena; when and under what circumstances they occur is still up for debate (Warren and Halpern-Manners 2012), and how to engage them remains a mystery. As Elder (2018:22) remarked about the children who participated in the Oakland Growth Study, "continued participation in the study may well have had some effect on the subjects, and there is no satisfactory way to assess this impact."

Furthermore, what we know comes almost exclusively from panel studies anchored in industrialized countries (e.g., 30 rounds of data from the German Panel Study, four waves of the National Study of Adolescent Health in the United States), where requests to participate

in research for political, scientific, and commercial purposes have been commonplace for more than a century (Chandon, Morwitz, and Reinartz 2004; Lazarsfeld 1940; Porst and Zeifang 1987). Despite the fact that massive amounts of survey data are being collected continuously across the globe, we know almost nothing about how the disruptive aspects of survey research play out in low-income contexts, where survey research is a more recent innovation (c. 1950s), and the introduction of new people (i.e., interviewers, enumerators, researchers), topics, and questions to a community may comprise a particularly unusual set of phenomena (Randall et al. 2013; Weinreb 2006).

Take, for example, the potentially disruptive influences of Demographic Surveillance Systems (DSS). First introduced in the 1960s, a DSS collects comprehensive data on demographic behavior, social change, and epidemiological trends in low-income settings where state-registered vital statistics are unavailable. Each DSS monitors a carefully defined geographic area, documenting every shift in birth rates, death rates, and migration. The typical survey interval within DSS sites is three to four months, and survey rounds often include collection of biological specimens, as well as detailed information on health and related aspects of culture (INDEPTH Network 2018). In 2017, 50 active DSS sites operating across Africa, Asia, and Oceania collectively released data on 1.6 million individuals (approximately 25 million person-years of observation) covering a wide range of topics, from causes of mortality to HIV/AIDS treatment effects to environmental change (INDEPTH Network 2016).

Many DSS sites have been used as settings for important clinical trials; valuable new methods like verbal autopsies have been developed in these research settings; and many health policies in the Global South have been formulated based on data collected from these sites, earning them the nickname “population laboratories” (D’souza 1981; Heuveline 2016). The importance of DSS sites to research, policy, and practice would be hard to overstate, but the impact of research practices in communities like these remains poorly understood. One recent study from a DSS site in Nepal candidly described how participants had become irritated by requests to participate in research and had begun demanding immediate, direct benefits (Aryal et al. 2012). But respondent fatigue is only one of many concerns that fall under our umbrella concept of disruption. Characterizing prolonged data collection efforts as a “passive intervention,” some scholars have raised concerns about the consequences of monitoring (Tollman and Zwi 2000; Ye et al. 2012). However, how best to go about systematically analyzing the levels of disruption that research inflicts on a population under surveillance remains an open question.

A Framework for Disruption

Weaving together insights from ethnography and survey methodology with our first-hand experience fielding a large-scale data collection effort brought us to *disruption*—an umbrella concept that encompasses all the ways research might influence people and populations. We offer a conceptual depiction of disruption in Figure 1, which we designed as a template for thinking about how research might influence respondents’ thoughts and behaviors—sometimes with consequences for their friends and neighbors. The curved arrows indicate the possibility that research effects can extend even beyond the population about which

inferences are being made. Importantly, we differentiate disruption from another key concern of survey research—error. Survey methodologists' insight that some employed respondents will answer "out of labor force" simply to shorten survey time (Halpern-Manners and Warren 2012) exemplifies why this distinction is necessary. When research effects influence only the responses, we consider it error.

In contrast to methodologists' focus on data and data quality, the concept of disruption refocuses our attention on the respondents themselves. At one end of the spectrum, disruption encompasses how the Framingham Heart Study changed participants' eating habits simply by asking them to fill out dietary inventories (Levy and Brink 2005). At the other end of the spectrum, disruption recognizes that the Kinsey reports were a watershed because they prompted conversations in living rooms across the country and transformed understandings of normal sexuality, with consequences for sexual behavior itself and for all subsequent research on sexuality (Igo 2008; Johnson-Hanks 2018).

Disruption is not necessarily negative; we use the term descriptively, not evaluatively, to index any kind of interruption to the normal course of events. Because observational research is fundamentally interactive, all research on human subjects should be thought of as susceptible to disruption. This is not to say that all research necessarily disrupts, but we venture that disruption can occur whether communities of interest are disadvantaged (e.g., Srinivas, Stack) or composed of highly educated professionals (e.g., Bosk). No populations should be assumed immune.

We concede that some aspects of disruption are more measurable than others, but we see no reason to constrain a framework for theorizing disruption by what is measurable in the short-run. Panel studies that are community-based may, of course, be most disruptive. But we posit that studies need not be longitudinal, large, intensive, or community-based to be disruptive. In the instance of the Kinsey reports, small and non-representative studies had a major impact on public opinion and, arguably, on sexual behavior in the United States—quite possibly even worldwide.

Approach

Our article proceeds in a slightly unorthodox fashion. First, we describe Balaka, Malawi—the setting in which we first began reflecting upon and then, later, empirically assessing the disruptive nature of survey research. Second, we offer "tales from the field," which depict disruption in Balaka and illustrate a few of its many pathways. Throughout the entirety of the study period (2009 to 2015), we accumulated a long and varied written record about our team's extra-survey interactions with respondents, including field notes written by investigators, project supervisors, and enumerators. The tales we present are drawn from this record.⁴ Third, we elaborate a quasi-experimental approach to identifying the scope and scale of disruption in longitudinal studies. We designed the refresher-sample-as-comparison (RSC) method in 2009 and implemented it in 2012 and 2015. Given our substantive foci, we leveraged the RSC to understand how our study shaped the way young adults thought about

⁴.From our perspective, it is unfortunate that data of this sort usually remain buried in personal records, diaries, field notes, and other kinds of marginalia.

and made decisions around relationships, sex, and childbearing. Fourth, we summarize data two ethnographers gathered in our research site during 2015 and use the ethnography to define more precisely the scope of disruption our research imparted in Balaka. Finally, we reflect on the empirical findings from our particular case and conclude with some general remarks about what acknowledging the disruptive nature of data collection means for scientific knowledge.

STUDY SETTING: BALAKA AND TLT

Our setting is Balaka Township, in the southern region of Malawi, where the Tsogolo La Thanzi (TLT) study was fielded between 2009 and 2015. TLT was designed to address a set of questions about the multi-directional relationships between relationship formation, fertility, and HIV risk in the context of a generalized AIDS epidemic. In April 2009, the research team conducted a complete household listing of all residences within a seven-kilometer radius of Balaka's main market; immediately thereafter, we drew a representative sample of 1,562 women between the ages of 15 and 25 and recruited them to participate in the panel study. The response rate at the baseline interview was 96 percent ($N = 1,505$) (Trinitapoli and Yeatman 2018; Yeatman et al. 2019). According to the coterminous round of the Demographic and Health Survey (NSO-Macro 2010), the TLT sample is similar in key sociodemographic variables to other young adults in the same age group in Malawi's southern region.⁵

Although typical of longitudinal survey research in many respects, TLT differs from other household-based surveys in four important ways. The first has to do with the structure of interviews: TLT interviews were conducted at a centrally located research center, in private rooms where responses could not be overheard by family members and neighbors.

Second, one substantive hallmark of the TLT study was the pairing of a focus on demographic processes with an interest in the subjective realm, including preferences, perceptions, and expectations. In other words, in addition to collecting information about respondents' characteristics and cataloging salient events that had taken place in their lives, TLT researchers asked a lot of questions about the future. Do you expect to get married in the coming year? Do you expect your current relationship to last? How long do you think you'll live? TLT focused on topics like sexual behavior, relationship satisfaction, worries about HIV risk, and testing for HIV—domains of life that are (1) subject to change during this stage of the life course, (2) highly sensitive and subject to normative pressures (i.e., sexual behavior and childbearing expectations), and (3) high-stakes (i.e., given limited access to treatment at the time, HIV infections are deadly, not inconvenient).

Third, in its initial phase, TLT collected follow-up rounds of survey data from all women at four-month intervals for a total of eight waves, spanning an almost-three-year period (2009 to 2011). At each wave, women were given an incentive (500 Malawian kwacha or 3 USD) for their participation, in accordance with Institutional Review Board guidelines. During this

⁵See Documentation Guide 11: Comparison of National-Level DHS Data to TLT Study (https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/5/707/files/2019/06/TLT_Documentation_Guide_11_DHS_Report.pdf).

period, the data collection effort was intensive, raising real questions about the impact of the research endeavor on respondents' lives. Study participants were prompted to think about their futures not just one time, but up to eight times, during precisely the period in which many of these events happen. (In Malawi, marriages, pregnancies, and new HIV infections are all concentrated during young adulthood.) For these reasons, prolonged engagement with a long-term study might represent a significant disruption—something beyond what might be engendered by an occasional request to participate in a census, political poll, or other household survey.

A fourth, unique characteristic relates to the overall study design: because both HIV infection and pregnancy are dyadic phenomena, we designed TLT with a focus on couple-level dynamics. During their first interview (and at all subsequent interviews), women who reported any ongoing romantic or sexual partnerships were given tokens with unique identifiers and asked to share these tokens with their male partners, who could then enroll themselves in the study. The tokens may serve as a particularly disruptive force and possible mechanism by which the study might influence union formation and dissolution, two of the processes we were most interested in studying.

SIGNS OF DISRUPTION: TALES FROM THE FIELD

We begin our analyses with tales from the field. These tales are distilled from a long, varied, and systematic record of qualitative data gathered in Balaka over the study period. Culturally attuned survey researchers who listen in their field sites, maintain a habit of field notes, and pay attention to the world outside the survey itself will have similar stories to tell about the rumblings and murmurs heard in their respective study settings. If researchers anywhere begin to look for evidence of disruption, they will almost certainly find it. To illustrate our own process, we adapted the general *spheres of disruption* framework (see Figure 1) as an analytic device for thinking through disruption in the particulars of our own study. Figure 2 first specifies the relevant spheres of sample, target population, and catchment area. Second, it locates the particular actors who emerge in these tales—they can be thought of as specific individuals and as general types of actors. Third, we use arrows to point to the likely processes through which our tales suggest that disruption occurs.

Tale One: Recruitment Catalyzes Conversations

The first tale was written in 2009 by one of our most experienced male interviewers during recruitment when interviewers visited respondents' homes, before our first interviews had even been conducted.

I came to a certain household which I would like to report here. There was a respondent at this household. When I asked about the name of the respondent, her mother confirmed that I was at the right house and that the girl was off chatting with her friend not far away. Her mother sent a daughter of the respondent's sister to call her and hurriedly come with her. The sky was dull, showering, and the mother offered me a chance to enter their house and sit on the sofa waiting for the respondent to come. During our waiting she began asking me more about our

project; she said it was almost three days that she had been seeing people moving around asking for people by name.

I took courage, telling her in detail as trained and as written in the consent form. Looking at the respondent's coversheet it showed that she was a minor (age 16), so I knew I had to present the consent form⁶ and get the mother to agree. I began telling her about the project and each of the points written on the assent form. She clearly understood, and I would guess that this mother had been to school. I told her that when her daughter comes back, I would also explain to her why I had come. I would ask her as a parent to sign on the form showing that she had also agreed that her daughter should participate in Tsogolo La Thanzi research. And the daughter would have to agree on her own. The mother said that the way she understood it, it seems that the project that has come to Balaka will be helpful.

Shortly later, the respondent came home. She knelt down near where I was and shook my hands to greet me. She then went to sit on the sofa adjacent to where her mother was sitting. Before I spoke, her mother told her that I was her visitor. I told her not to be afraid of anything. I began explaining why I was there. She agreed and confirmed her name and the year she was born. It was correct. I gave her the assent form that her mother had signed and then asked for a witness to sign down. After I signed in the middle, the respondent's sister went to call her. She came and signed. Then I gave her the calendar, circling the appointment date, and her mother assured me that she would make sure to keep it safe. She also said that she was going to be the one to remind her when the day comes. I thanked her.

Then she went on, saying: "This is good, because maybe these young girls are going to change their bad behavior of going around with their boyfriends. They are very young, but instead of completing their education and working to get a good job, they are just playing and having fun with boys." She said "this one especially" (meaning the respondent). "Please try to advise her there so that maybe she will finally understand the behavior of going with boyfriends. I advise her, but it is not working. Maybe she thinks that we are of the same age."⁷

I cut her words [interrupted] saying, "No, mother! Please understand. This project is designed for research, but not to teach the youth new ways." They could keep the assent form and read more to thoroughly understand on their own, when I left, what the project was all about.

The respondent did not say anything. Her sister was also quiet and pretended to be reading the form I had given them.

The mother continued saying that she was very worried of how the world is today. She said everywhere people are greatly worried about "ndimatenda opanda mankhwalawa" which translates to "this disease that has no medicine." I agreed. And she continued saying: "It's very painful when the youth are dying with AIDS,

⁶In cases where respondents were under the age of 18, consent was given by the parents and assent by the respondent herself.

⁷Elders advise and their advice is almost always taken seriously; age mates in Malawi, on the other hand, are not authority figures.

leaving us old people behind. Who will develop the country? Who will take care of their brothers and sisters? How can we depend on these youth?” She said she is very worried because she has born six children, all of which are female. She continued that she wished she had a son because it gives you confidence when you have educated a son that he will go far with education. She said a daughter may drop out of school at any time due to pregnancy when you as a parent have spent a lot on her education. I agreed by nodding but adding no words. Rain showers stopped and then I thanked them for securing me in their house from showers. She said, “Don’t worry we are all Malawians. If you want to wait a bit longer, you can.” She said she was going to tell the daughters to help each other prepare nsima⁸ fast. I thanked her for the offer but told them not to worry about me.

Then I left this beautiful house, leaving this smart, plump woman, a lorry on the ground. And she said she is wishing the best to the Tsogolo La Thanzi Project. She said in English, “I am wishing your project all the best!” I thanked her and then proceeded with my recruiting.

Without relaying any particularly noteworthy event, this field note conveys something subtle and methodologically important. Before the TLT surveys even began, the introduction of our research project sparked conversations in homes. Who are the TLT people? What are they studying? Why are they testing blood? What does the inside of that building look like? How do I get a job there? How do I get into that study? During recruitment season, these types of conversations were widespread across the catchment area and were even occasionally reported to us from neighboring towns. Throughout the recruitment process, to spread accurate information and correct misinformation (e.g., that we might be collecting and selling blood), the primary tools we used were information sessions with local leaders (particularly religious leaders and village chiefs), town-hall-style meetings held in villages across the catchment area, and our recruitment scripts in households.

This exchange between a young woman and her mother, with her sister present, was distinct from other chatter in that it related directly to our substantive foci. This conversation was certainly not a first—the mother in this story seems to have been nagging her daughter for some time. But it raised for us the possibility that TLT may have brought new actors into these conversations, increased their frequency, or altered their scope—all with unknown consequences for related attitudes and behaviors.⁹ It would be unwise for us to make any statements about prevalence from our field notes, but we can confidently conclude a few things from our longer running record on disruption: (1) This interaction was not unusual: our interviewers documented similar interactions between parents and children (minor respondents), female respondents and their sisters, and husbands and wives. (2) The conversations our study stimulated were normatively laden (i.e., the perils of “bad

⁸.Nsima is maize porridge, the staple Malawian food.

⁹.In Tale 1, the new actors in the mother/daughter exchange are the sister and the interviewer; in other stories collected across the community during fieldwork, conversations about TLT involved brief encounters between vendors at the market and other shoppers or between a respondent and the bike-taxi driver providing her transport. We posit changes in frequency and scope of conversations as a possibility rather than as a fact. To this end, we draw heavily on Watkins’s arguments about the salience of informal conversational networks for demographic outcomes and the special relevance of clinics, markets, and public spaces for these conversations (Ashforth and Watkins 2015; Kohler, Behrman, and Watkins. 2001; Rutenburg and Watkins 1997; Watkins and Danzi 1995; Watkins and Swidler 2009).

behavior”) in ways that the carefully worded questions we asked in the survey were not. In other words, the conversations we identified as arcing toward disruption were topically resonant with our study aims and qualitatively distinct from the merely clarifying chatter.

Tale Two: Incentives Introduce Inequalities

Trinitapoli’s field note from Wave 1 interviews, July 2009:

A father comes to the research center to have a look around. His daughter had been here, he said. TLT is a new thing, and its people unknown, so he wanted to see for himself that the place was okay and make sure there wasn’t anything untoward going on in this building. He didn’t seem upset or worried—more curious than anything else. A handful of fathers and uncles came to see us prior to the start of data collection and during the first week....

After chatting for a bit, this father had a complaint. “My second-born is participating in your study. But, you see, I have three daughters. You can see my problem?” Abdul, our project manager, explained that we selected the respondents “at random using a computer” and that for the purposes of science, only those who had been selected could participate. We couldn’t add other women to the study because enrollment had to work like a lottery to meet scientific standards. He gave a great explanation.

But the daughter had come home with her incentive (500 kwacha – 3 USD). She announced that the money had been given to her, not to the family, and that she would be getting the same amount at each subsequent visit. The father wanted clarity from us about whether the money was for the girl to spend as she liked or whether it was intended to help the whole family.

This tale illustrates a more tangible way through which randomly allocated opportunities to participate in research affect respondents—by putting money in their pockets. Compensation levels (cash and gifts) are fiercely regulated by relevant Institutional Review Boards (Emanuel 2005) and disclosed in methods sections as a matter of standard practice. The politics, ethics, and consequences of compensating respondents for their participation in scientific research has been widely examined, usually with an emphasis on the consequences for response rates and data quality.¹⁰ Little is known, however, about how monetary incentives influence respondents and their households. As this father pointed out, even small amounts, allocated at random, may generate new inequalities within households and across the study setting.

Tale(s) Three: Tokens Reach Partners and Partnerships

As described earlier, TLT was designed to collect couple-level data in addition to individual-level data. To accomplish this, we asked women about their sexual and romantic relationships and then gave them tokens they could pass along to the men they talked about,

¹⁰In the only study we know of to have tested for resentment effects, in a study community in Mexico, Stecklov, Weinreb, and Winters (2016) find that exclusion from randomized controlled trials may seed resentment and frustration within a community, with deleterious consequences for response rates. Evidence from India suggests the converse is also true: incentives enhance response rates among beneficiaries, yet sometimes at the expense of data quality (Stecklov, Weinreb, and Carletto 2017).

so these men could also enroll in the study. As the study progressed, we began to suspect that the exchange of tokens may be affecting, indeed disrupting, romantic relationships.

During Wave 3, one of our female interviewers recounted the following story to the last author, who wrote it in her field notes:

The interviewer needed to travel from the research center in town to one of the more distant villages to follow-up with a respondent who had recently given birth and was, by consequence, unable to come into the research center herself. Our interviewer hires a bicycle taxi to make the trek and ends up having a long conversation with the driver, who is friendly and asks her a lot of questions about her work. When the bike-taxi driver shares that he is participating in our study, and has a funny story to tell, the interviewer replies: “Well, I love a funny story, and we have a long drive ahead of us. Please tell me more!”

Back in November, his girlfriend had completed her second interview at the center. The two of them were chatting about their days, and she described her visit to TLT and says, “Mmmm. They ask a lot of questions, and they also want to test you for pregnancy.” He knew of some people who had gotten tokens from their girlfriends and wives. So he asks her: “Well, where’s my token?” And she says, “Listen: I didn’t tell them about you as my partner. You only have half of your stuff at my house. I don’t know where the rest of your things are, and I don’t know how serious you are.” He says, “I love you and I thought we were all-but-married. But if you don’t even want to tell the TLT people about me then maybe you’re ashamed of me, and *you* are the one who is not serious.” So they agree [over this issue of tokens] that this is, indeed, a real relationship. And they agree that at the next interview, she will tell the TLT people about him, get a token, deliver it to him, and get him into the study too.

In contrast to this tale, in which the token seems to have concretized the relationship, other tales around tokens indicated they were playing a role in relationship dissolution. One respondent traveled a long distance (to a neighboring village) to bring a token to the man she had reported as her fiancé, only to find another woman there cooking nsima for him. This act of cooking was not just about food; it was a clear indication of householding—the man had deceptively presented himself to the respondent as single. Another respondent reportedly caught her fiancé red-handed while delivering a token, creating a full-blown spectacle in his village; this story circulated within Balaka for weeks and was recounted to us by four different staff members who had heard the gossip around town. In her subsequent interview, the respondent herself told her interviewer exactly what had unfolded—anticipating a set of questions about her love life and changes since her last interview, she did so in advance of the relationships “section” of the questionnaire. In yet another case, a respondent reported having gotten divorced between the time of her last interview and this one. Capturing the relevant details in the “marriage” module of the questionnaire, the interviewer checked the box to indicate “other reason” and noted in the associated blank: “Husband did not like a visit to Tsogolo La Thanzi, so that after visiting here he shouts at her and ends the marriage packing and went away for good. Since then he no longer came back.” Over the course of

the study, we heard almost a dozen other stories about tokens, making it difficult to guess the “net effect” of tokens on relationships.

The relationships depicted in the last two stories may have had short shelf-lives to begin with. But the exchange of tokens seemed to have kicked off a process of discovery, revealing new insights about some of these romantic partners. Whether leading to an intensification of the relationship or to conflict and relationship dissolution, the tokens themselves were consequential and, at least occasionally, disruptive.

As the TLT study progressed, our desire to learn about disruption systematically rather than anecdotally increased. Was the complaining father unusually sensitive about intra-household equality, or was he telling us about something bigger that was changing as a result of TLT’s presence in the community? How often was this exchange of tokens a momentous event for couples? How much extra-survey conversation about sex, relationships, HIV, and imaginations of the future was prompted by our research activities? How far might the disruptive tremors of a community-based survey extend? The evidence we accumulated in these tales helped structure our approach to making disruption an analytic priority.

Returning to Figure 2, our tales from the field are just one of three methodological tools we leveraged to perceive and estimate disruption. We now turn to two other tools we added to our original protocols: the RSC to estimate disruption in the sample, and study-focused ethnography to gauge the scope of disruption in the community broadly. All three tools are stamped onto Figure 2, arranged to represent the specific aspects of disruption each can illuminate.

THE REFRESHER-SAMPLE-AS-COMPARISON APPROACH

To assess the impact of an intensive, longitudinal survey on respondents’ lives, we introduced the RSC. Back in 2009, we anticipated concerns with panel conditioning in our sample and pulled a second, random sample from the original household listing, which we then set aside. Given our growing concerns with disruption, rather than just panel conditioning over time, we decided to use this sample as a comparator to measure disruption in 2012. Conveniently, when the project was extended into 2015, we could also leverage the comparison sample as a refresher to compensate for attrition.

Figure 3 depicts the TLT study design and illustrates the key analytic comparisons facilitated by the approach. For the RSC to work as intended, refresher-sample women must be statistically comparable to core-sample women in terms of their age, family structure, and economic conditions at Time 1 (2009). This condition is met: at the time of the original household listing in 2009, the sampled groups ($N = 1,562$, $N = 527$) were statistically identical based on these sociodemographic characteristics, just as we would expect them to be based on random sampling procedures (analyses not shown, available upon request).

To assess whether these two groups *remain* comparable three years later (Time 2) given differential response rates, we use paired *t*-tests to compare the 2009 household-listing characteristics of the refresher sample we interviewed ($N = 315$) to those same measures for the core sample we retained to Wave 8 ($N = 1,212$). Summarized in Table 1, these

descriptive comparisons provide evidence that despite sizable differences in response rates for the two samples, on all key household- and individual-level measures, the core-sample respondents we followed all the way to 2011 are statistically identical to the refresher sample we found in 2012.¹¹ Subsequently, we treat Wave 8 (October to December 2011) and the first observation from the refresher-sample women (February to April 2012) as a single point in time. We compare responses from the 1,212 core-sample women and the 315 refresher-sample women across an array of outcomes, using various regression specifications to test for significant differences. Core-sample women are considered “treated” by participation in the study (i.e., interviewed up to eight times in three years) and refresher-sample women “untreated.”

During recruitment efforts for the follow-up study in 2015 (July to September), no distinction was made between women in the core sample ($N = 1,200$) and those in the refresher sample ($N = 255$).¹² To assess panel conditioning on a longer time horizon, this third comparison (Time 3) tells us whether the study effects detected in 2012 were durable (i.e., persistent three years later) or short-term and subject to decay by 2015.

Outcome variables.

We examine a wide range of outcomes, from concrete indicators of socioeconomic status to behavioral and attitudinal factors. Throughout the study period, respondents were repeatedly questioned about HIV/AIDS, sexual behavior, relationship quality, and fertility preferences (see Appendix Table A1 for descriptive statistics for all outcome variables). We expect the effects of research to manifest most poignantly in precisely the domains of life the study itself forces respondents to think about, evaluate, and respond to.

Key independent variable.

Across these analyses, our key independent variable is sample type (0 = refresher sample, 1 = core sample), which estimates the effect of having participated in the intensive portion of the TLT study.¹³

Controls.

Given that RSC design is quasi-experimental and treatment is “random” (despite the potential threat of non-response bias, confirmed by part 1 of our analyses, below), an extensive set of controls is unnecessary.

¹¹We note one attrition-related difference of possible consequence: as the study progressed, the core sample itself became slightly biased toward the disadvantaged, at least in terms of household infrastructure (improved roof). However, the same process seems present in the refresher sample when we compare the refresher sample we drew to the refresher sample we found. We see no analytic consequences of the difference for our task of assessing study effects, as we are comparing like-with-like both in 2011/12 and in 2015. Additional details are provided in Appendix Table A2.

¹²Two respondents from the refresher sample started, but did not complete, the 2015 survey and are excluded from our analyses, reducing our analytic N in 2015 by two cases (analytic $N = 1,453$).

¹³In some ancillary analyses, we looked to identify dose-response effects within the sample of TLT respondents, using number of waves contributed as a proxy for intensity of study participation. Because this issue is so intertwined with migration and retention, and because preliminary analyses did not suggest this was a promising avenue, we present here only the main difference between the two samples.

Refresher-Sample-as-Comparison Results

Table 2 presents results from more than a dozen regression models (logit, ordered/multinomial logit, poisson, or OLS, depending on the nature of the outcome). Each coefficient indicates the measured difference between samples (core versus refresher). The first column displays results from the pooled 2011/2012 observations, and the second column depicts longer-term effects—differences detected during the 2015 follow-up. The outcomes are grouped into categories, and our discussion of these categories begins with the most concrete: economic factors (because it is perhaps easiest to envision how money received from participation could influence respondents' socioeconomic status). We then proceed with our analyses of study-related phenomena: planning for the future, attitudes, behaviors, and outcomes related to HIV and childbearing (the study's most heavily-covered topics), and, finally, to a set of more distal outcomes.

Socioeconomic Indicators

As noted earlier, women received a small incentive for study participation at each wave, which would have amounted to almost 30 USD for core-sample women who participated in all waves. If the concerned father from Tale 2 was telling us something that resonated beyond the particular dynamics of his own household, we should find statistical evidence of a socioeconomic gap between core- and refresher-sample women over time.

Although core-sample and refresher-sample women lived under similar conditions in terms of household infrastructure in 2009 (Table 1) and 2012 (Table 2), by 2015, women who had participated in the study from its beginning were living in better homes.¹⁴ We see a similar pattern for ownership of personal items (e.g., watch and cell phone) and household goods.¹⁵ The same pattern holds true for educational attainment. In 2012, core-sample women had achieved roughly .6 more years of schooling than the refresher sample; by 2015, the difference in the highest level completed was almost one full year, a consequential difference in a setting where average schooling is approximately seven years. Over the course of the study, core-sample women came to have better household infrastructure, possess more personal items, own more household goods, and complete more school.

Planning

Given TLT's emphasis on expectations (around relationships and childbearing in particular), eight rounds of study participation demanded of respondents a non-negligible amount of attention toward the future and an articulation of their related thoughts. We measure future-orientation using the question: "How often, if at all, do you think about or plan for your future?" Responses range from never (0) to very often (4). Table 2 shows that at both points in time, core-sample women thought more about the future than did refresher-sample women.

¹⁴.In 2012 and 2015, we can assess differences along other lines, such as an improved floor (i.e., not a dirt floor) and latrine access. Overall, the results for infrastructure differences, assessed with a more diverse set of measures, compiled into an index using principal component analysis, are consistent with what we find from using only the roof as our indicator: negligible differences in 2012, with significant differences emerging in 2015.

¹⁵.See Appendix Table A3 for details about constitutive items and their coding.

We explore this pattern in more detail in Figure 4, which depicts the shifting distribution of our future-orientation indicator at each wave. At baseline, about 40 percent of core-sample respondents reported thinking about and planning for the future “very often,” and that proportion increased (nonlinearly, but steadily) as the study progressed. By the time Wave 8 was fielded, close to 60 percent of respondents said they were thinking about the future very often. When visited again in 2015, the proportion of core-sample women who thought a lot about the future had returned to 2009 levels, suggesting the increase observed between 2009 and 2011 may have been driven by the repeated questions we asked about the future. When we first spoke with refresher-sample respondents (represented by the dashed outline) in 2012, 45 percent reported thinking about their future very often; however, that number dropped significantly (to about 30 percent) in 2015. And thinking rarely/never about the future was twice as prevalent in the refresher sample (13 percent) as it was among core-sample respondents (6 percent).¹⁶

Study-Related Norms: Attitudes Related to Childbearing and HIV

If study participation prompted TLT respondents to think more about their futures than they otherwise would have, some of this thinking should manifest in their attitudes—particularly around the subjects TLT asked a great many questions about: childbearing and HIV. With respect to norms around childbearing, core- and refresher-sample women reported an ideal family size of around three at both points in time.¹⁷ Yet we find evidence that our study may have influenced how people think about fertility in a more general way. In 2015, core-sample women were far less likely than refresher-sample women to agree with the statement, “You don’t plan having children, they just happen” (43 versus 58 percent). This likely reflects the fact that these women were repeatedly questioned about their fertility intentions and plans throughout the study period and asked to reflect on the types of families they desired and *how many, when, and with whom* to have these children. A sizable majority of core-sample women believed that childbearing is within one’s control, but the far less agentic view that “children ... just happen” prevailed among refresher-sample women.

HIV was another focal point of the study. After having been primed to think about HIV as a general phenomenon (i.e., attitudes) and as it relates to their own life (i.e., subjective risk), by the time of the 2015 follow-up, core-sample women were significantly less likely to agree with the statements “HIV is God’s lottery” and “The changes I need to make to prevent HIV are a lot to ask.” This suggests the sustained conversation about HIV that core-sample women had with interviewers over the course of the study may have eroded notions that prevention is outside of one’s control.¹⁸

¹⁶To ensure the patterns observed here are not being driven by maturation, we conducted robustness checks using an age-standardized (21 to 25) subsample at all waves. These analyses produced substantively identical results.

¹⁷This is consistent with an earlier finding by Yeatman, Sennott, and Culpepper (2013), who tested for panel conditioning influences in the TLT data and found no evidence of a relationship between survey experience and ideal family size.

¹⁸We also examined, by sample, a host of attitudes and behaviors unrelated to our study: things like use of mosquito nets, religious practice, and gender roles. Across these outcomes we found no consistent differences between samples. For the sake of space, these analyses are not presented here but are available upon request.

Sexual Behavior and Contraception

Women who participated in the study from the beginning were indistinguishable from refresher-sample women in terms of their lifetime number of sexual partners and current contraceptive use.

Marital Status and Relationship Quality

Romantic relationships are the node by which sex, HIV, and childbearing are connected and a key factor through which many study effects we observe could operate. In 2012, core-sample women were more likely than refresher-sample women either to have never married or to have moved beyond their first marriage (e.g., divorced, separated, remarried). In other words, study participation seems to be associated with two departures from the typical marital trajectory that characterizes Balaka: (1) delayed entry into first marriage (where median age at first marriage is 18) and (2) accelerated dissolution of first unions (10 percentage points). This resonates with expectations we might have generated based on the token-related tales. Iterative and detailed conversations about expectations around marriage and relationship satisfaction prior to marriage, combined with the financial opportunity to stay in school longer, led some core-sample women to delay marriage and others to make difficult decisions around union dissolution. However, these differences in marital status between the two samples became nonsignificant by 2015, suggesting processes of *delay* and *acceleration* rather than a manifestation of *difference*.

Beyond these differences in marital status alone, the RSC provides some evidence that study participants are in better relationships. Although they were less likely to agree with the general statement that “these days, most married men are faithful to their wives,” core-sample women were less likely to suspect their own partner of having other partners. In 2015, they reported higher levels of satisfaction in their current relationship. In other words, core-sample women were less optimistic about men’s faithfulness generally, but they were also less suspicious of the particular men with whom they were in relationships, and they reported higher levels of relationship quality with those men. Higher-quality relationships may tie back to the socio-economic advantages we observed among core-sample women, which may have had less to do with the incentives themselves than with having achieved better relationships with better men.

Study-Related Outcomes

We consider two study-related outcomes that could be considered end-points. From HIV biomarkers collected in 2012 and 2015, we find no statistical difference in HIV prevalence among core- and refresher-sample women. Although the proportion of women without children was the same between samples at both points in time, total parity among core-sample women was slightly lower than parity of their refresher-sample counterparts at both points. By September 2011, a date that accounts for the small lag in the refresher sample interview, the average difference between samples was $-.12$. The parity difference persists into 2015, when core- and refresher-sample women were interviewed simultaneously. An array of study effects, including delayed marriage, higher educational attainment, and repeated contemplation of childbearing and future planning manifest here in either the

postponement of higher-order births or adoption of other family-limitation strategies (but not contraception) among core-sample women.

General Health and Well-Being

The word “disruption” often connotes negative effects, but this need not necessarily be the implication. In fact, we read the cultivation of more agentive attitudes, future-oriented thinking, and smaller families as disruptions because they are departures from what is typical in Balaka, not because they are negative. Our notion of disruption does, however, imply that research has consequences for the very topics and phenomena under investigation. Significant differences between core-sample and refresher-sample women in the domain of health and well-being are strong and, we believe, indicative of research consequences that extend beyond the behaviors under observation.

At both points in time, core-sample women reported better health. Moreover, core-sample women scored far better than refresher-sample women on a simple mental-health scale (range 0 to 4, composed of questions like “I have felt lonely” and “I have felt depressed”). In 2012, study participants reported higher levels of overall life satisfaction, but that difference did not persist into 2015.¹⁹

Repeated questions about health, combined with the collection and notification of biomarkers (pregnancy and HIV) throughout the study may have primed core-sample women to think more intentionally about their health and well-being and perhaps behave in health-enhancing ways (Edwards 2018). Moreover, the simple act of listening can be a strong force, with real, if small, effects on physical and mental health. In low-resource settings where women’s voices are often ignored, the simple act of asking questions and recording answers can give women a sense of self-worth. This insight was clearly stated in Van den Borne’s (2005) ethnographic study of Malawi, in which she observed that her interviews made her respondents *feel significant*. Just as a radio reporter interviews famous people—and records their answers to get everything right—Van den Borne’s respondents felt acknowledged and affirmed as they shared their life stories, and they said that being heard felt good to them.

Social Desirability Bias

As noted in the general framework of disruption (Figure 1), study effects sometimes influence responses (error) without generating any real changes in respondents themselves (disruption). To explore the extent to which study participation may have simply taught respondents to present themselves in a positive light, we combined six agree/disagree statements into a simple additive index to capture social desirability bias. The index is composed of statements like “I always share food with hungry people” and “I am sometimes rude.” These items have clear moral underpinnings that almost all women in Balaka would aspire to but that are next to impossible to actually achieve.²⁰ The “correct” answer in each statement is coded as 1, so larger numbers indicate a tendency toward social desirability

¹⁹Because our research setting is partly agricultural and characterized by known epidemiological fluctuations by season, we restricted our analyses in the health domain to outcomes that are not susceptible to seasonality.

²⁰See Appendix Table A3 for details about constitutive items and their coding.

bias. This index allows us to see whether study participants gave more socially desirable answers as the study progressed or, alternatively, became more honest over time. On average, core-sample women scored higher on the social desirability index in 2015 than they did in 2009 (4.12 and 3.50, respectively). However, we find no difference in the social desirability of answers given by the core and refresher samples in 2015, which suggests aging (not study participation) explains this increase within the core sample over time.

Overall, our analysis shows that study participation brought changes to our respondents' lives. Big and small, these changes in attitudes and behaviors were not limited to the scope of research. Survey participation within TLT influenced an array of outcomes: first, by providing small but continuous monetary incentives to the respondents, and second, by propagating future-oriented thinking and spurring numerous conversations about relationship quality, sexual behavior, and HIV-related norms. Moreover, research elicited emotional and behavioral responses among core-sample women, resulting in improved health at both points in time. Our results clearly support the assertion that researcher-participant interactivity in social science research is a concrete phenomenon with real consequences. Because the interactivity of research often manifests in unexpected and unintended ways, good social science needs more contemplation of and open discussion about the "pitfalls and surprises" researchers encounter in the research process and the ad hoc decisions they make in response (Leahey 2008).

ETHNOGRAPHIC SYNTHESIS FROM 2015

Returning to Figure 2, the RSC is located at the boundary between the sample of respondents who participated in the study and the population they were supposed to represent. The RSC approach cannot, however, tell us anything about how TLT may have influenced the broader community. If research is affecting the entire community, this may not be estimable at all, but would lead us to conclude that the RSC estimates are a conservative appraisal of the study's effect—as those in the refresher sample would have felt the study's presence before their first interview. Epistemologically, the notion of disruption is troubling because it violates a key assumption of the pre-post perspective to estimating change in the social world. But as stated earlier, just because something cannot be measured well does not mean we should not try to assess it. Acknowledging that there may, indeed, be no way to empirically estimate disruption across a study setting or beyond it, we used study-focused ethnography to gauge the scope of disruption and illuminate the processes by which it is manifest.

In 2015, TLT incorporated two dedicated ethnographers into the study. One European woman and one Malawian woman (E and M, respectively) lived together in a village at the outskirts of the TLT catchment area for approximately five months, after which time M moved to town, where she lived for three months—overlapping with our peak months of data collection. E and M lived in these communities as researchers interested in understanding young adulthood, childbearing, relationships, and access to health care. Neither woman advertised her relationship to TLT. Because both women had lived in the village for a full year back in 2009 for an unrelated ethnographic study, their presence in the village in 2015 was not seen as particularly odd or in any way connected to TLT.

In addition to documenting daily life and conversations occurring in the village and in town, we asked the ethnographers to catalog all conversations they overheard or engaged in about TLT. In response, E and M captured dozens of new actors and accounts we characterize as simultaneously uneventful and note-worthy. Looking to the edges of the framework depicted in Figure 2, E and M picked up relevant conversations among local vendors who expected to see an increase in their sales. The owner of a stationary store planned regular visits to TLT to ensure we would buy pens, envelopes, and other sundry office materials from him rather than sourcing from the big cities; and the town pharmacist started stocking extra inventory of rubber gloves and band-aids, which TLT personnel were buying frequently: “TLT is our big customer, they always buy a lot from us compared to other people.”

On a few occasions, groups of unemployed men sat together, speculating about TLT salaries, discussing whether to apply for a job as a watchman, and wondering how its current staff went about securing their positions at the research center. Market vendors remarked on the extraordinary beauty of the women (interviewers and respondents) who could be found coming and going from the TLT research center—bathed, lotioned, and nicely dressed, as if for a special occasion. Overall, the ethnographic record collected in 2015 suggests chatter and gossip about TLT was widespread but innocuous: Why are they back? Are any jobs available? Look at the girls coming and going. Can I profit from this venture? In the nearly 100 references to TLT that E and M documented over the eight-month period, we found no strong evidence that our study was shifting social interactions in any profound way or introducing new kinds of information or materials into everyday life.

Nonetheless, some Balaka residents observed that TLT seemed to be having a positive effect on respondents. The incentives were a major topic of conversation; respondents who had been drawn into our sample “by lottery” were thought of, by their non-participating peers and elders, as the beneficiaries of a unique and fortunate opportunity. In August 2015, after one respondent completed her final of nine interviews, her distraught mother came to an interviewer’s residence to ask if she had been dishonorably “dismissed” from the study. Concerned that her daughter may have “failed TLT exams,” she pleaded with the interviewer to please allow her daughter to complete more interviews into the future. From her perspective, TLT encouraged good behavior and the incentive helped offset expenses like school uniforms, soap, and other incidentals that supported her schooling.

While reviewing these data in real-time in 2015, it seemed implausible to us that this mother’s claim about the role of TLT in her daughter’s “good behavior” could be empirically confirmed on the aggregate. But through the ethnographic data E and M collected, we learned that many residents shared this perception (and we heard no stories suggesting the converse). Older community members remarked that TLT was reducing teen pregnancy in Balaka. Conversations to this end were documented at project town-hall meetings (when local project staff reported back on study findings) and appeared as a recurring theme in TLT-related chatter we read in both sets of field notes. In addition to comments from parents and everyday remarks overheard at the market and bus stops, this specific insight about reduced fertility was made by chiefs, religious leaders, teachers (including one secondary-school headmaster), and nurses.

Overall, ethnographic data gathered in 2015 lend strong support to the qualitative evidence we accumulated as tales during phase 1 of the study, and they also conform quite closely to the conclusions we began to draw from our analyses of the RSC. Throughout Balaka, TLT was thought to have influenced respondents' lives in positive ways: enhancing health and economic well-being, bolstering relationships, and helping women manage their fertility.

Although it would be impossible to rule-out ripple effects completely, the ethnographic effort turned up no evidence of widespread disruption throughout the catchment area, and it tended to point back to the boundary between the studied and the unstudied as the key line of disruption in TLT. Taking an earthquake as a metaphor, the main shock of disruption is located among core-sample women, with aftershocks affecting the larger community. Whereas the RSC is a more precise tool for estimating intensity of disruption from a very particular location, study-focused ethnography works like a seismogram, helping us locate the hypocenter of disruption and its relative magnitude along the seismic wave. Alone, neither tool can generate a full assessment of the magnitude, scope, origin, and intensity of a study's impact. But together, the RSC, interviewer fieldnotes, and study-focused ethnography generate strong evidence about how and to what extent disruption operates.

DISCUSSION

"In what ways is a society changed by the very tools employed to represent it?" – Igo (2008:4)

This article advances two major contributions. First, we make an empirical contribution to understanding the extent to which data collection disrupts the phenomena we intend to observe. The second goal of our study is more general—to raise the disruptive nature of research as a likelihood instead of a mere possibility and suggest that rather than destroying our ability to know anything, serious engagement with disruption will improve knowledge and generate productive, new insights about the problems we set out to understand.

While running a longitudinal study of young adulthood in Malawi, we channeled our nagging concerns about disruption into two, complementary research strategies designed to clarify how regular research practices disrupt both the lives of study participants and the larger community. The data we collected using qualitative and ethnographic techniques suggest that research alters social life by stimulating corners of the local economy, generating new questions, altering expectations, prompting reflections about the future, and igniting new conversations. By adding a refresher sample to the TLT study and leveraging it analytically as a comparison, we were able to estimate the magnitude of associated changes in respondents' characteristics, attitudes, and behaviors. We find ample evidence, across study-related domains and beyond, that these reflections led to conscious and subconscious changes—changes big (leaving a spouse) and small (buying a new pair of shoes).

The findings from our *in situ* experiment are simultaneously disquieting and comforting. We find that research participation has a tangible effect on respondents, and the differences are overwhelmingly positive for respondents' lives. Four of these disruptions are especially salient to the context of young adulthood in Malawi. First, study participation had a durable effect on socioeconomic status: boosting consumption in the short run, increasing household

wealth on a longer time-horizon, and yielding higher levels of educational attainment among respondents. Second, core-sample women were more likely to have postponed or exited a first marriage, and their current relationships were of higher quality than those experienced by refresher-sample women. Repeated conversations about relationship quality prompted reflection that may have catalyzed impulses to move on from “low quality” relationships. Third, core-sample women consistently reported better self-rated health and lower incidence of depression and loneliness. Perhaps most importantly, by repeatedly speaking about their future and uncertain issues around HIV status and fertility, core-sample women were thinking more about the future and tended to believe that HIV and childbearing were within their control. Although the measured parity difference between samples is open to debate and interpretation, we think it is reasonable to conclude that the study itself, without intending to, averted over 100 births to core-sample women.²¹ The conversations our ethnographers cataloged in town and in villages in 2015 suggest that at least some of these changes were visible to community members and attributed to TLT.

In a study about HIV, childbearing, and relationships, we unintentionally ignited a process of contemplation about the future that was especially consequential to the domains of life our study engaged. If asking about the future is a self-fulfilling prophecy, then the consequences of simply having asked are non-negligible. For example, if you ask a person “Are you thinking about an elephant?” they might say no, but now they are thinking about one, if only a little bit.²² How powerful, exactly, are these prompts to think about the future in specific ways? To be listened to? To be told that your answers matter uniquely and need to be recorded for scientific purposes?

Although we can see no reason to think that only young adults would be susceptible to the influences of research, we concede that disruption may be more resonant among populations that have been historically disadvantaged—those for whom “being heard” or being represented in science may be new. To this end, we heard from a colleague that the original cohort of the Mature Women in the National Longitudinal Studies (NLS) may have also exhibited disruption:

They were first interviewed in 1967, a time when most research was done on men; they were 30–44 at first interview, and the retention rates remained quite high until the last round, fielded in 2003. The interviewers often reported how seriously these women took their role as respondents; they felt that they were “standing in” for millions of other women about whom we knew very little.²³ In early waves, when women were asked at what age they planned to retire, large proportions answered either “don’t know” or “when my husband does.” They were also asked about pensions and savings and investments. By the next wave, the proportion of women who answered by stating an actual age had grown. We imagined the kitchen-table conversations and the possible planning that had been initiated by that question. We

²¹Point estimate of $-12 \times 1,200$ women = -144 births. The 95 percent confidence interval is quite wide and suggests considerable uncertainty (between 36 and 252 births averted).

²²Martin (2016; see also Martin and Frye 2012) refers to this as the “refrigerator light” problem.

²³It was not until almost 20 years later that Senator Pat Schroeder was able to get a bill enacted that required researchers to explain why women should be excluded from study populations.

thought it served like a prompt: If I'm being asked about it, it's important for me to think about and have a good answer for the next time. The early survey literature on attitudes contemplated a similar process, wondering the extent to which being asked one's opinions served as a prompt to formulate some, one hopes on the basis of a bit of research. All this is to say that data collection is a social interaction, with information being exchanged in both directions, and the interaction itself signaling a level of interest, and perhaps value, that many respondents may not be used to. (private communication, Melissa Hardy, November 6, 2018)

In sharing the hidden underbelly of disruption in TLT and the lack of precise control scholars have over the research process, the second purpose of our study is more general—to acknowledge the disruptive nature of research and recommend that social scientists embrace disruption as a fact of life and begin laboring to understand it, rather than ignoring or concealing it. As evidence of disruption accumulates from across a wide range of research settings, we would not be surprised to learn that small-area and longitudinal studies are most susceptible to disruptive forces. Researchers working in these traditions, in particular, would be wise to weigh the risks of disruption against the many benefits of closely-spaced surveys (see also Zwane et al. 2011).

A wider knowledge base will also be critical for questioning long-standing habits in experimental research. Because the decision to collect a baseline has real consequences for research findings, knowledge about disruption will put a finger on that cost-benefit scale, too. Across geographic settings and methodological traditions, it must become the researcher's responsibility to actively contemplate the extent to which research activities *might* disrupt individuals and communities, and then incorporate into research plans systematic procedures that can be leveraged to document disruptions and assess their magnitude.

From a study-design perspective, the benefits of using a refresher sample to proxy disruption is clear. RSC is a simple methodological tool that can be used to answer specific and general questions about how research activities matter. It gives researchers involved in original, longitudinal data collection efforts a tangible and (relatively) inexpensive tool. Retrospective approaches may also be generative, particularly in cases where insights are derived from record linkages, and the unstudied can be matched to, for example, death certificates, tax returns, or medical records.

However, all data-gatherers and data-users should think more about the role of disruption in their work and how they might address it. Experiments, team-based ethnography, panel-conditioning models, field notes, and cognitive interviewing may all have a role in establishing the kind of foundation we would like to build our science on. Expanding the arsenal of tools—observational and quasi-experimental—with which we can detect disruption, identify its scope and magnitude, and speak to its implications for our claims fits squarely within the movement to elevate rigor and transparency in social science research.

Replication and transparency standards in the social sciences have evolved considerably in recent years and will remain a topic of fundamental import for the foreseeable future. Many publishers now routinely require statements about conflicts of interest, IRB approvals, data

accessibility, and the sharing of code for replication. In our view, voluntary *disruption disclosures* would be a natural and positive extension of this trend. Given the methodological heterodoxy that characterizes sociology, this type of research transparency may be an especially fitting and fruitful avenue for our field. We can imagine both boring and arresting disruption disclosures being delivered as footnotes, codas, prefaces, robustness checks, appendices, or as main text.

How would our field change if all scholars working on human populations took steps to make disruption effects more transparent by testing for it statistically, describing it ethnographically, and unveiling many of the affirming stories we have heard from colleagues around the globe while presenting this research as in-progress work? We venture that if sociology were to move in this direction, we could quickly develop a grasp of the distribution of disruption in sociological research—some situations in which disruption levels destabilize initial results, many in which disruption is nonexistent or negligible, and others in which disruption leads to completely new questions and research agendas. However, until a critical mass of scholars coordinate (however loosely) to address disruption explicitly in their written work, it will be impossible for us to know whether disruption of the type we identified through TLT is particular to young adults, to impoverished contexts, to the domains of sexual and reproductive health, or is indexing a far more general phenomena.

Schwartz and Jacobs (1979) noted that the notion of “influencing” a study setting is relative, and it is up to researchers’ discretion to decide what kind of influence they will allow themselves to exert on the scene. More recently, Igo (2008:4) wrote, “Surveys are a peculiar sort of social investigation in which the public is simultaneously object, participant, and audience.” Survey researchers, in particular, have to decide how much disruption their models can tolerate, whether disruption is correctable (statistically), or when, in extreme cases, disruption means we can no longer credibly generalize from a sample to the unstudied population it was carefully designed to represent. In extreme cases, disruption levels may indicate that an observational study should be halted. In TLT, we detected evidence of disruption in the domains of economic life, marriage, future-oriented thinking, study-related attitudes, health, and well-being. The differences are consistent, significant, and small. In other words, our research did not transform respondents’ lives completely, but did lead to small changes that may have been meaningful.

Once we puncture the assumption of non-interactivity in survey research, we must confront the fact that the data we rely on are the product of negotiations between interviewers and respondents. This raises challenging questions about what to do next. Is it even possible to simply observe the social world? Or does any attempt to do so systematically alter the world being observed, complicating the project of making inferences to non-studied populations? We agree with Martin (2016) that we cannot avoid using interactive techniques to gather data about the vast majority of questions that animate the social sciences. Researchers can, however, implement systematic approaches like the RSC and study-focused ethnography to learn more. We can interrogate our data to find out more about the kind of residue data-generating endeavors like TLT necessarily leave behind. And rather than dreading the

results, we can get excited about the possibilities and opportunities new insights about disruption will almost certainly bring.

Acknowledgments

We thank John Levi Martin and Marco Garrido for helpful comments on an early draft, and Rob Warren for motivating us to pursue this line of research back in 2009. In-progress iterations of this article have been presented at the Minnesota Population Center, the Guttmacher Institute's Research Brownbag Series, Penn State University's Population Research Institute, the sociology departments at Peking University and Zhejiang University, and the University of Chicago's Social Theory and Evidence Workshop.

Data note

This article uses data from Tsogolo La Thanzi, a research project designed by Jenny Trinitapoli and Sara Yeatman and funded by grants R01-HD058366 and R01-HD077873 from the National Institute of Child Health and Human Development. TLT data are available through restricted data-user agreements managed by Data Sharing for Demographic Research (DSDR) at the University of Michigan.

APPENDIX

Table A1.

Descriptive Statistics for All Outcome Variables

	CS (2011) Mean	RS (2012) Mean	CS (2015) Mean	RS (2015) Mean	Min.	Max.
Finished roof	.43	.41	.53	.46	.00	1.00
Personal items	1.58	1.43	1.68	1.46	.00	4.00
Household goods	2.55	2.44	2.70	2.37	.00	8.00
Pair of shoes owned	3.04	2.60	3.49	3.16	.00	40.00
Years of education	8.09	7.43	8.23	7.33	.00	15.00
Think about future	3.38	3.02	3.03	2.70	.00	4.00
Ideal family size	3.37	3.47	3.53	3.70	1.00	8.00
Children just happen			.43	.58	.00	1.00
HIV is God's lottery			.14	.23	.00	1.00
Changes to prevent HIV			.67	.81	.00	1.00
Lifetime N of sexual partners	1.69	1.65	1.66	1.67	.00	20.00
Currently using condom	.14	.13	.15	.15	.00	1.00
Other modern contraceptives	.36	.40	.53	.56	.00	1.00
Never married	.30	.27	.16	.13	.00	1.00
First marriage	.47	.57	.56	.56	.00	1.00
Beyond first marriage	.23	.16	.27	.30	.00	1.00
Married men are faithful			.30	.37	.00	1.00
Satisfied with relationship			3.90	3.82	1.00	4.00
Suspect current partner	.13	.88	.31	.41	.00	1.00
HIV positive	.10	.08	.15	.17	.00	1.00
Nulliparous	.31	.26	.14	.11	.00	1.00
Number of children	1.33	1.64	2.11	2.38	.00	6.00
Self-rated health	3.86	3.75	3.84	3.64	1.00	5.00
Mental health	3.55	3.41	3.33	3.22	.00	4.00
Satisfied with life	4.35	4.14	4.46	4.35	1.00	5.00

	CS (2011) Mean	RS (2012) Mean	CS (2015) Mean	RS (2015) Mean	Min.	Max.
Social desirability index			4.12	4.13	.00	6.00
<i>N</i>	1,212	315	1,200	253		

Source: TLT Wave 8, refresher-sample baseline dataset, and TLT-2.

Table A2.

Supplementary Analyses to Assess Attrition

	Core (2011–2009)	Core (2015–2009)	Core (2015–2011)	Refresher (2015–2012)
Ethnicity				
Yao	.00	.02	.00	–.02
Chewa	–.01	.01	.00	.00
Lomwe	–.01	.01	.00	.00
Ngoni	.02	.02	.00	.02
Other tribe	–.01	.00	.00	.00
Year of birth	–.11	.13	.06	–.01
Native to Balaka	.04 *	.02 *	.01	–.02
Household size	–.02	.09	.05	.00
Improved roof	–.05 *	.02 *	.01	–.03
Any children prior to 2009	.03	.02	–.02	.01
Poor health	.00	.00	.00	.00

Source: TLT household listing, core-sample baseline dataset, core-sample Wave 8 dataset, refresher-sample dataset, and TLT-2.

Note: Mean differences (i.e., simple subtraction) are presented in each column. We used *t*-tests to assess differences within the core and refresher samples over time.

*
 $p < .05$

**
 $p < .01$

 $p < .001$ (two-tailed tests).

Table A3.

Coding Details for All Outcome Variables

Model	Outcome Variables	Coding
1. economic Variables		
logit	<i>Finished roof</i>	What type of roof does your house have? 1 = asbestos, cement, iron sheets 0 = grass thatch
poisson	<i>Personal items</i>	Do you personally own: (1) wrist watch (2) mobile phone (3) a pair of jeans (4) luggage
poisson	<i>Household goods index</i>	Does your household own a working:

Model	Outcome Variables	Coding
		(1) bed with mattress
		(2) television
		(3) radio
		(4) land line or mobile phone
		(5) refrigerator
		(6) bicycle
		(7) motorcycle
		(8) car or truck
poisson	<i>Pair of shoes owned</i>	How many pairs of shoes do you personally own?
OLS	<i>Years of education</i>	What was the last standard/form that you completed?
2. planning		
ologit	<i>Think about future</i>	How often, if at all, do you think about or plan for your future?
		0 = never
		1 = almost never
		2 = sometimes
		3 = fairly often
		4 = very often
3. Attitudes related to Childbearing and HIV		
poisson	<i>Ideal family size</i>	People often do not have exactly the same number of children they want to have. If you could have exactly the number of children you want, how many children would you like to have?
logit	<i>Children just happen</i>	You don't plan having children, children just happen.
		1 = agree
		0 = disagree
logit	<i>HIV is God's lottery</i>	HIV is God's lottery.
		1 = agree
		0 = disagree
logit	<i>Changes to prevent HIV</i>	Changes I need to make to prevent HIV are a lot to ask.
		1 = agree
		0 = disagree
4. Sexual Behaviors and Contraception		
poisson	<i>Lifetime N of sexual partners</i>	How many people overall have you ever had sex with in your lifetime?
logit	<i>Currently using condom</i>	(For ongoing partnership) Think about the last three times you had sex with [NAME], did you use condoms?
		1 = every time, two times, once
		0 = never
logit	<i>Other modern contraceptives</i>	Apart from condoms, are you currently using contraception with [NAME]?
		1 = pill, injectable, loop, sterilized, or norplant
		0 = no methods, string, traditional medicine
5. marital Status and relationship Quality		
Marital status is constructed using two questions:		
1. Your current marital status is?		
2. How many times have you been married in your lifetime, including your current marriage?		

Model	Outcome Variables	Coding
mlogit	<i>Never married</i>	Current marital status is NEVER MARRIED
	<i>First marriage</i>	Current marital status is MARRIED and this is the first marriage
	<i>Beyond first marriage</i>	All others
logit	<i>Married men are faithful</i>	Most married men are faithful to their wives 1 = agree 0 = disagree
ologit	<i>Satisfied with current relationship</i>	In general, how satisfied are you with this relationship? 1 = very dissatisfied 2 = somewhat dissatisfied 3 = somewhat satisfied 4 = very satisfied
logit	<i>Suspect current partner (2011/12)</i>	Do you suspect or know that [NAME] has had sexual relations with women apart from you since you have been together? 1 = yes-know, suspect 0 = can't know what he does, probably not
logit	<i>Suspect current partner (2015)</i>	About how many sexual partners do you suspect or know [NAME] has had in the past four months NOT including you? 1 = more than one suspect 0 = no suspect
6. Study-related Outcomes		
logit	<i>HIV positive (from biomarker)</i>	1 = HIV positive 0 = HIV negative
logit	<i>Nulliparous</i>	How many children have you ever given birth to, including ones that may have died? 1 = no child 0 = one or more children
poisson	<i>N of children ever born</i>	How many children have you ever given birth to, including ones that may have died?
7. general Health and Well-Being		
ologit	<i>Self-rated health^a</i>	In general, would you say your health now is: 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor
Four questions assessing mental health were recoded as dichotomous variables and summed to construct a mental health scale: 1 = very often, fairly often, and sometimes; 0 = almost never and never.		
ologit	<i>Mental health</i>	(1) I have felt depressed. (2) I have felt life was not worth living. (3) I have felt content. ^a (4) I have felt lonely.
ologit	<i>Satisfied with life^a</i>	How satisfied are you with your life, all things considered? 1 = very satisfied 2 = satisfied

Model	Outcome Variables	Coding
		3 = somewhat satisfied
		4 = somewhat unsatisfied
		5 = very unsatisfied
8. Social Desirability Bias		
Six agree/disagree questions were used to construct an additive measure of social desirability bias index. Each question is coded 1 if agree, 0 if disagree.		
OLS	<i>Social desirability index</i>	(1) I am sometimes rude. ^a
		(2) I go to every development activity in my village.
		(3) I always follow the advice of doctors and nurses.
		(4) I always share food with others during hunger seasons.
		(5) I sometimes gossip about others. ^a
		(6) I go to every funeral in my village.

^aReverse coded for analysis.

Biography

Jeong Hyun Oh is a PhD student at the University of Chicago studying demography and development. Her research focuses on the new and persistent dimensions of inequality in sub-Saharan Africa, including school-to-work transitions, health disparities, subjective social status, and the consequences of these for demographic behaviors.

Sara Yeatman is Associate Professor of Health and Behavioral Sciences at the University of Colorado Denver. Her research examines the causes and consequences of HIV and unintended fertility in sub-Saharan Africa and the United States.

Jenny Trinitapoli is Associate Professor of Sociology at the University of Chicago. She has written extensively about the role of religion in the AIDS epidemic in sub-Saharan Africa, and since 2008 has been the principal investigator of Tsogolo La Thanzi.

References

- Adler Patricia A., and Adler Peter. 1987 *Membership Roles in Field Research* Newbury Park, CA: Sage Publications.
- Aryal Umesh Raj, Vaidya Abhinav, Shakya-Vaidya Suraj, Petzold Max, and Krettek Alexandra. 2012 "Establishing a Health Demographic Surveillance Site in Bhaktapur District, Nepal: Initial Experiences and Findings." *BMC Research Notes* 5:489. [PubMed: 22950751]
- Ashforth Adam, and Watkins Susan. 2015 "Narratives of Death in Rural Malawi in the Times of AIDS." *Africa* 85(2):245–68.
- Barber Jennifer S., Gatny Heather H., Kusunoki Yasamin, and Schulz Paul. 2016 "Effects of Intensive Longitudinal Data Collection on Pregnancy and Contraceptive Use." *International Journal of Social Research Methodology* 19(2):205–22. [PubMed: 26783387]
- Binswanger Johannes, Schunk Daniel, and Toepoel Vera. 2013 "Panel Conditioning in Difficult Attitudinal Questions." *Public Opinion Quarterly* 77(3):783–97.
- Bosk Charles L. 2003 *Forgive and Remember: Managing Medical Failure* Chicago: University of Chicago Press.
- Chandon Pierre, Morwitz Vicki G., and Reinartz Werner J.. 2004 "The Short- and Long-Term Effects of Measuring Intent to Purchase." *Journal of Consumer Research* 31(3):566–72.

- Christakis Nicholas A., and Fowler James H.. 2013 “Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior.” *Statistics in Medicine* 32(4):556–77. [PubMed: 22711416]
- D’souza Stan. 1981 “A Population Laboratory for Studying Disease Processes and Mortality: The Demographic Surveillance System, Matlab Comilla, Bangladesh.” *Rural Demography* 8(1):29–51. [PubMed: 12338517]
- Edwards Ryan D. 2018 “If My Blood Pressure Is High, Do I Take It to Heart? Behavioral Effects of Biomarker Collection in the Health and Retirement Study.” *Demography* 55(2):403–34. [PubMed: 29520638]
- Elder Glen H. 2018 *Children of the Great Depression* Boulder, CO: Westview Press.
- Emanuel Ezekiel J. 2005 “Undue Inducement: Nonsense on Stilts?” *American Journal of Bioethics* 5(5):9–13. [PubMed: 16179296]
- Halpern-Manners Andrew, and John Robert Warren. 2012 “Panel Conditioning in Longitudinal Studies: Evidence from Labor Force Items in the Current Population Survey.” *Demography* 49(4):1499–519. [PubMed: 22893185]
- Halpern-Manners Andrew, John Robert Warren, and Florencia Torche. 2014 “Panel Conditioning in a Longitudinal Study of Illicit Behaviors.” *Public Opinion Quarterly* 78(3):565–90.
- Halpern-Manners Andrew, John Robert Warren, and Florencia Torche. 2017 “Panel Conditioning in the General Social Survey.” *Sociological Methods & Research* 46(1):103–24. [PubMed: 28025587]
- Hardy Melissa. 2018, 11 6 Personal communication
- Heuveline Patrick. 2016 “The Mekong Island Population Laboratory (MIPopLab), A Demographic Surveillance System in Rural Cambodia (2000–06).” *Data Sharing for Demographic Research* (<https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/36601>).
- Igo Sarah E. 2008 *The Averaged American: Surveys, Citizens, and the Making of the Public* Boston: Harvard University Press.
- INDEPTH Network. 2016 “Annual Report 2016” Accra, Ghana: INDEPTH Network Retrieved June 19, 2018 (<http://www.indepth-network.org/reports/annual-report-2016>).
- INDEPTH Network. 2018 “About Us” Accra, Ghana: INDEPTH Network Retrieved June 19, 2018 (<http://www.indepth-network.org/about-us>).
- Johnson-Hanks Jennifer. 2018 “Distributing Norms.” Presented at University of Chicago, April, Chicago, IL.
- Kohler Hans-Peter, Behrman Jere B., and Watkins Susan C.. 2001 “The Density of Social Network and Fertility Decisions: Evidence from South Nyanza District, Kenya.” *Demography* 38(1):43–58. [PubMed: 11227844]
- Kreuter Frauke, Presser Stanley, and Tourangeau Roger. 2008 “Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity.” *Public Opinion Quarterly* 72(5):847–65.
- Kroh Martin, Winter Florin, and Schupp Jürgen. 2016 “Using Person-Fit Measures to Assess the Impact of Panel Conditioning on Reliability.” *Public Opinion Quarterly* 80(4):914–42.
- Lazarsfeld Paul F. 1940 “‘Panel’ Studies.” *The Public Opinion Quarterly* 4(1):122–8.
- Leahey Erin. 2008 “Methodological Memes and Mores: Toward a Sociology of Social Research.” *Annual Review of Sociology* 34:33–53.
- Levav Jonathan, and Fitzsimons Gavan J.. 2006 “When Questions Change Behavior: The Role of Ease of Representation.” *Psychological Science* 17(3):207–13. [PubMed: 16507060]
- Levy Daniel, and Brink Susan. 2005 *A Change of Heart: How the People of Framingham, Massachusetts, Helped Unravel the Mysteries of Cardiovascular Diseases* New York: Vintage Books.
- Martin John Levi. 2016 “Toward a Nightmare-Resistant Sociology.” *Contemporary Sociology* 45(5):535–42.
- Martin John Levi, and Frye Margaret. 2012 “A Dialogue between Theory Section Award Winners.” *Perspectives: Newsletter of the ASA Theory Section*, 12 Retrieved June 19, 2018 (<http://home.uchicago.edu/~jlmartin/Debate/Conversation%20with%20Frye.pdf>).

- Moorman Robert H., and Podsakoff Philip M.. 1992 "A Meta-Analytic Review and Empirical Test of the Potential Confounding Effects of Social Desirability Response Sets in Organizational Behavior Research." *Journal of Occupational & Organizational Psychology* 65(2):131–49.
- NSO-Macro. 2010 Malawi Demographic and Health Survey 2010 Calverton, MD: ORC Macro.
- Paulhus Delroy L. 1984 "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology* 46(3):598–609.
- Porst Rolf, and Zeifang Klaus. 1987 "A Description of the German General Social Survey Test-Retest Study and a Report on the Stabilities of Sociodemographic Variables." *Sociological Methods & Research* 15(3):177–218.
- Randall Sara, Coast Ernestina, Compaore Natacha, and Antoine Philippe. 2013 "The Power of the Interviewer." *Demographic Research* 28(27):763–92.
- Rutenberg Naomi, and Susan Cotts Watkins. 1997 "The Buzz Outside the Clinics: Conversations and Contraception in Nyanza Province, Kenya." *Studies in Family Planning* 28(4):290–307. [PubMed: 9431650]
- Schwartz Howard, and Jacobs Jerry. 1979 *Qualitative Sociology: A Method to the Madness* New York: Free Press.
- Srinivas Mysore Narasimhachar. 1976 *The Remembered Village* Berkeley: University of California Press.
- Stack Carol B. 2008 *All Our Kin: Strategies for Survival in A Black Community* New York: Basic Books.
- Stecklov Guy, Weinreb Alexander, and Carletto Calogero. 2017 "Can Incentives Improve Survey Data Quality in Developing Countries? Results from a Field Experiment in India." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(4):1033–56.
- Stecklov Guy, Weinreb Alexander, and Winters Paul. 2016 "The Exclusion from Welfare Benefits: Resentment and Survey Attrition in a Randomized Controlled Trial in Mexico." *Social Science Research* 60:100–109. [PubMed: 27712671]
- Struminskaya Bella. 2016 "Respondent Conditioning in Online Panel Surveys: Results of Two Field Experiments." *Social Science Computer Review* 34(1):95–115.
- Tollman Stephen M., and Zwi AB. 2000 "Health System Reform and the Role of Field Sites Based upon Demographic and Health Surveillance." *Bulletin of the World Health Organization* 78(1):125–34. [PubMed: 10686747]
- Torche Florencia, John Robert Warren, Andrew Halpern-Manners, and Valenzuela Eduardo. 2012 "Panel Conditioning in a Longitudinal Study of Adolescents' Substance Use: Evidence from an Experiment." *Social Forces* 90(3):891–918.
- Tourangeau Roger, and Smith Tom W.. 1996 "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60(2):275–304.
- Trinitapoli Jenny, and Yeatman Sara. 2018 "Tsogolo La Thanzi (TLT): Baseline Wave, Malawi, 2009–2012 [Healthy Futures]" Ann Arbor, MI: Interuniversity Consortium for Political and Social Research [distributor], 2018-10-22 (10.3886/ICPSR36863.v3).
- Van den Borne Francine. 2005 *Trying to Survive in Times of Poverty and AIDS* Amsterdam: Het Spinhuis Publishers.
- Watkins Susan Cotts, and Danzi Angela D.. 1995 "Women's Gossip and Social Change: Childbirth and Fertility Control among Italian and Jewish Women in the United States, 1920–1940." *Gender & Society* 9(4):469–90.
- Watkins Susan Cotts, and Swidler Ann. 2009 "Hearsay Ethnography: Conversational Journals as a Method for Studying Culture in Action." *Poetics* 37(2):162–84. [PubMed: 20161457]
- Warren John Robert, and Andrew Halpern-Manners. 2012 "Panel Conditioning in Longitudinal Social Science Surveys." *Sociological Methods & Research* 41(4):491–534.
- Weinreb Alexander A. 2006 "The Limitations of Stranger-Interviewers in Rural Kenya." *American Sociological Review* 71(6):1014–39.
- Williams Patti, Block Lauren G., and Fitzsimons Gavan J.. 2006 "Simply Asking Questions about Health Behaviors Increases Both Healthy and Unhealthy Behaviors." *Social Influence* 1(2):117–27.

- Wooden Mark, and Li Ning. 2014 "Panel Conditioning and Subjective Well-Being." *Social Indicators Research* 117(1):235–55.
- Ye Yazoume, Wamukoya Marilyn, Ezech Alex, Emina Jacques B. O., and Sankoh Osman. 2012 "Health and Demographic Surveillance Systems: A Step towards Full Civil Registration and Vital Statistics System in Sub-Sahara Africa?" *BMC Public Health* 12(1):741. [PubMed: 22950896]
- Yeatman Sara, Chilungo Abdallah, Lungu Sydney, Manadingo Hazel, and Trinitapoli Jenny. 2019 "Tsogolo La Thanzi: A Longitudinal Study of Young Adults Living in Malawi's AIDS Epidemic." *Studies in Family Planning* 50(1):71–84. [PubMed: 30690738]
- Yeatman Sara, Sennott Christie, and Culpepper Steven. 2013 "Young Women's Dynamic Family Size Preferences in the Context of Transitioning Fertility." *Demography* 50(5):1715–37. [PubMed: 23619999]
- Zwane Alix Peterson, Zinman Jonathan, Van Dusen Eric, Pariente William, Null Clair, Miguel Edward, Kremer Michael, Karlan Dean S., Hornbeck Richard, Giné Xavier, Duflo Esther, Devoto Florencia, Crepon Bruno, and Banerjee Abhijit. 2011 "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108(5):1821–6.

Figure 1.
A General Framework of Disruption in Social Science Research

Figure 2.

Application of General Disruption Framework to the Specific Characteristics of TLT

Note: The particular characters are represented in person figures, arrows show the processes of disruption, and the methodological tools leveraged to perceive/estimate disruption are in italics.

Figure 3.
Timeline of Tsogolo La Thanzi Data Collection

Note: This figure depicts the timeline of the TLT from 2009 to 2015, with specific attention to the response rate over time for core and refresher samples. *N*s appear in boxes as stand-alone numbers and response rates (with original sampled women as the denominator) are shown in parentheses.

Figure 4.

Distribution of Time Spent Thinking about Future by Wave and Sample

Source: TLT Waves 1 through 8, refresher-sample baseline dataset, and TLT-2.

Note: We indicate core-sample observations with a solid outline and distributions for the refresher sample with a dashed outline.

Retrospective Descriptive Overview of Core- and Refresher-Sample Characteristics from the Time of the Household Listing (2009), Balaka, Malawi

Table 1.

	Core Sample	Refresher Sample
Native to Balaka ^a	.65	.67
Age at listing	19.60	19.74
Household size	5.20	4.98
Never married at listing	.50	.49
Any children prior to listing ^a	.47	.51
Poor health	.02	.03
Finished roof	.40	.36
<i>N</i>	1,212	315

Sources: Tsogolo La Thanzi (TLT) household listing (2009), core-sample baseline dataset (2009), and refresher-sample baseline dataset (2012).

Note: With the exception of age, all measures are dichotomous. Any children born prior to the household listing were retrospectively constructed from detailed birth histories collected at baseline. Poor health is a dichotomous measure indicating that the respondent was described as “ill” or “very ill” by their household’s informant at the time of the listing. We used *t*-tests to test for significant differences between core- and refresher-sample women retained in 2011/2 and found no differences.

^aVariable was constructed from the baseline questionnaire.

Table 2.

Core- and Refresher-Sample Differences at Two Points in Time

Dependent Variables	2011/2 Comparison	2015 Comparison
Socioeconomic Indicators		
Finished roof	.11	.29 *
Personal items	.10	.14 *
Household goods index	.04	.13 **
Pair of shoes owned	.16 ***	.10 **
Years of education	.58 **	.90 ***
Planning		
Think about future	.66 ***	.51 ***
Attitudes related to Childbearing and HIV		
Ideal family size (IFS)	-.03	-.05
You don't plan having children, children just happen		-.61 ***
HIV is God's lottery		-.59 **
Changes to make to prevent HIV are a lot to ask		-.73 ***
Sexual Behaviors and Contraception		
Lifetime N of sexual partners	.03	-.00
Currently using condom ^a	-.13	-.03
Other modern contraceptives ^a	-.14	-.12
Marital Status and relationship Quality		
Never married	.29 *	.20
First marriage	(base)	(base)
Beyond first marriage	.53 **	-.10
Most married men are faithful to their wives		-.34 *
Satisfied with current relationship ^a		.54 *
Suspect current partner of having other partner ^a	-3.98 ***	-.42 **
Study-related Outcomes		

Dependent Variables	2011/2 Comparison	2015 Comparison
HIV positive	.28	-.20
Nulliparous	.33	.28
Number of children ever born	-.21***	-.12**
General Health and Well-Being		
Self-rated health	.22*	.41**
Mental health	.38**	.26*
Satisfied with life	.49***	.22
Social Desirability Bias		
Social desirability index		-.00
N	1,527	1,453

Source: TLT Wave 8, refresher-sample dataset, and TLT-2.

^aQuestion asked only of women in an ongoing sexual relationship (2011/2 comparison = 1,154; 2015 comparison = 1,145).

*
 $p < .05$

**
 $p < .01$

 $p < .001$ (two-tailed tests).