BMJ Open

Network Spillover Effects and Follow-Up Correlates in a HIV Prevention Intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759
Article Type:	Research
Date Submitted by the Author:	21-Aug-2019
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Network Spillover Effects and Follow-Up Correlates in a HIV Prevention Intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]*Perelman School of Medicine, University of Pennsylvania* [b]*Center for Health Equity Research and Promotion, Philadelphia VA* [c]*Department of Global Health and Social Medicine, Harvard Medical School* [d]*Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine* [e]*Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences* [f]*Management and Development for Health* [g]*Department of Population Health, London School of Hygiene and Tropical Medicine* [h]*Department of Biostatistics, Harvard T. H. Chan School of Public Health* [i]*University of Texas School of Public Health*
[j]*Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu*

Preprint submitted to BMJ Open *August 21, 2019*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spill over into these network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64) and HIV knowledge (OR=20.0); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17). The 14.2% increase in NMs' HIV knowledge was largely due merely to CAs participating in the intervention, rather than CAs transmitting *new* knowledge to their NMs.

Conclusions: Characteristics of social network members of PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, indicating that intervening in highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population not easily accessible via other means.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar trials may spillover in a population.
- Although the high dropout rate of social network members meant we could examine factors leading to dropout, it exposed the outcomes to bias if those who dropped out were different from those who did not.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the YLL due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practice and injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that low adherence will compromise their effectiveness [4]. However, neither method is fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, many trials have been conducted in persons living with HIV (PLH) which have increased HIV knowledge as an endpoint and have often found positive effects of interventions on HIV knowledge [7, 8].

Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect is a change to an individ-

ual's behavior that is due to an exogenous change to another's. In the context of an intervention, it means a change to a non-participant in the intervention stemming from someone else's participation in the intervention. What remains unknown, however, is whether or not a spillover effect exists for HIV knowledge during and after an intervention.

New knowledge can come from a variety of sources, one of the most important of which is a person's social network [10, 11]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [12, 13, 14]. This has been directly examined in participant-driven interventions, where early participants directly educate members of their social network [15]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [16]. Work on diffusion through social networks has shown that spreading intervention effects beyond the initial population increases the cost-effectiveness of these interventions [17]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions.

Additionally, the networks of PLH are often hidden networks [18], due to the continued stigma of HIV and AIDS in many settings [19]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective [20]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects [21]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who enroll in the intervention itself is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conducted a study on network members of PLH enrolled in the larger *NAMWEZA* intervention [22]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [23].

Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors transfer will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

3.1. Study Population

We analyzed social network data from the Agents of Change trial [24], which was a stepped-wedge randomized controlled trial that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices [25]. CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania. Participants were given a baseline questionnaire and were randomized to one of three waves in which to receive the intervention. The intervention comprised 10 weekly structured sessions aimed at empowering PLH to become HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [26]. Within one month of each wave of the intervention, CAs and NMs were given follow-up surveys. The study lasted from November 2010 to January 2014.

At baseline, participants were also asked to recruit members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. These network members could be either HIV positive or negative. Upon successful recruitment, these network members were also given a baseline survey. The NM was only aware the CA was a participant in the intervention if the CA shared this information with them, which many did not due to HIV-related stigma [27]. Each CA therefore formed a CAN-NM pair with each NM they recruited. Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed a follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyas.

3.2. HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [28]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other

characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex practices to reduce the risk of transmission. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa [29], whereas it was 0.78 in the original study. It has also been translated to Swahili, with only minor differential item functioning [30]. This indicates that the measure performs adequately in other, similar populations, relevant to this work. Because the majority of CAs and NMs answered all questions correctly, we summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

3.3. Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in temrs of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs could have more than one NM, we assessed significance using univariate intercept-only Generalized Estimating Equations (GEE) with a Normal or Bernoulli distribution for continuous and dichotomous variables, respectively. We used exchangeable working correlation structures in these models. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing all their follow-up interviews, we fit a logistic regression to determine predictors of follow-up. In order to examine the association between these same variables and the time to follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM completed their baseline interview. The outcome was loss-to-follow-up. NMs were interval-censored between the date of their last completed interview and last available date of the next scheduled interview, as NMs could have decided not to continue to participate at any point during that time.

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [23], we aimed to elucidate the putative causal mechanism through which the CAs impacted their network members. As shown by **(author?)** [31], social network spillover effects can be broken down into direct and indirect effects in the case of dyadic relationships (Figure 1). This method has since been used for novel spillover analyses [32]. Although previous studies showed that this same type of analysis cannot

be done on full network data, the data in this study consisted of only paired individuals, the CAs and their NMs, so in this case the analysis does not result in biased estimates [33, 34].

The method parses social influence into direct and indirect effects.The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [15]. The natural direct effect (NDE) is the effect merely receiving an intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if all CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [31]. The data are not publicly available due to the sensitive nature of HIV infection status.

3.4. Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants. Participants assessed the burden of the intervention via qualitative interviews; we found that many felt the timing was burdensome.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, p¡0.0001), but were less likely to have at least 7 years of education (52.0% vs. 52.3%,

p¡0.0001). Only 12.3% of NMs were HIV-positive, compared to *all* CAs (p<0.0001). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.0001
Female	380 (53.7%)	349 (53.9%)	0.956
Employed	490 (69.3%)	356 (55.0%)	<0.0001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.0001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.78
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.0001
Married	373 (52.7%)	338 (51.1%)	0.61
HIV Positive	87 (12.3%)	662 (100%)	<0.0001
Private source of water	309 (43.7%)	263 (39.7%)	0.20

Table 1: Demographic characteristics at baseline, with the results of an intercept-only GEE for differences.

Logistic regression showed that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female, having complete HIV knowledge, being employed, and being married were all significant predictors of increased odds of being followed-up. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (OR=0.81, 95% CI: 0.67,0.98) as did the NM living with HIV (OR=0.42, 95% CI: 0.18,0.99). CA having a private water source was significantly associated with increased odds of the NM being followed-up (OR=2.17, 95% CI: 1.33,3.57), even after controlling for the NM having a private source of water.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married and the NM living with HIV significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to treatment via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [23]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.0060,0.02), which is the effect the intervention had via the CA's HIV knowledge changing, and a natural direct effect (NDE) of 12.3% (95%

Characteristic (N=459)	Adjusted OR (95% CI)	Adjusted HR (95% CI)
Characteristics of NMs		
Gender (Female)	1.64* (1.02,2.63)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	20* (3.70,125)	2.20 (0.97,5.01)
Employed	1.5* (1.04,2.5)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.81* (0.67,0.98)	0.92 (0.83,1.01)
Married	1.72* (1.04,2.86)	1.28* (1.01,1.64)
Living with HIV	0.42* (0.18,0.99)	0.71* (0.51,0.99)
Having a private source of water	0.86 (0.54,1.37)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.17* (1.33,3.57)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)

Table 2: Results of multivariate logistic regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM was followed-up, and the continuous outcome of time-to-follow-up, respectively. "*" indicates significant at the $p < 0.05$ level.

CI: 0.06,0.19), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge.

5. Discussion

Based on results from the *NAMWEZA* trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that as much information as possible is transmitted from the study participants to members of their social networks.

We found many significant differences between CAs and the NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that the NMs were being recruited from populations that were substantially different for the population from which the CAs came [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their

social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample [35]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [36]. This means that future studies either need to focus recruitment on employed persons, or encourage spillover from study participants to network members.

Although the study design potentially accessed a separate slice of the population than other methods vis-á-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. Had CAs and NMs been more similar, and any of these characteristics acted in a causal fashion, the CAs may have had greater influence over their NMs, resulting in greater follow-up of the NMs due to encouragement from their CAs. We also note that less HIV knowledge and living with HIV predicted greater loss to follow-up among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE - i.e. knowledge spilled-over as a result of the CAs participating in *NAMWEZA*, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because HIV knowledge of all CAs started off relatively high (80%), so increases in HIV knowledge based on our questionnaire was not possible for all CAs. Following the intervention, CAs' average HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was likely merely their association with the CA. So instead of the CA providing informational support by directly sharing novel information [37], CAs may have felt more empowered to share their existing knowledge as a result of the intervention.

This finding is important for future interventions: spillover effects of this intervention will likely carry over only one degree of separation from the CAs (opposed to spreading indefinitely in a snowball effect), as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs

receive. Because NMs do not actually receive the intervention, but only the contact with their CA, it is unlikely that they would feel empowered to become CAs themselves, thus limiting the spread of the intervention. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who are highly connected in the community would maximize the potential number of links by which spillover can occur. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver the intervention to others, the effects seen here could spillover continuously.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our final analysis may have biased our results. Even though the exposure was randomized, the loss to follow-up can result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Second, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Third, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways.

6. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Understanding

how this then translates into closeness of ties between NMs and CAs and the implications for follow-up may inform strategies for retention and may increase the impact of future HIV behavioral interventions. The mediation analysis presents a compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. The results presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

7. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

8. Contributorship

JL, MCSF, and JPO developed the research plan. JL conducted the analysis and wrote the draft of the manuscript. YL assisted in verification of analyses. All authors discussed the results and contributed to the final manuscript.

9. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available.

10. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[11] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[12] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[13] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[14] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[15] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[16] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[17] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[18] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[19] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[20] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[21] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[22] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[23] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[24] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[25] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[26] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[27] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[28] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[29] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[30] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[31] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013 jun;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[32] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[33] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR. 2012;40(2):211–39.

[34] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

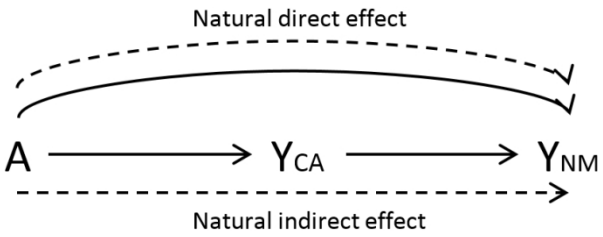
[35] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[36] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

[37] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

11. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE).



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Network Spillover Effects and Follow-Up Correlates in a HIV Prevention Intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759.R1
Article Type:	Original research
Date Submitted by the Author:	02-Jan-2020
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Global health, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Network Spillover Effects and Follow-Up Correlates in a HIV Prevention Intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]*Perelman School of Medicine, University of Pennsylvania* [b]*Center for Health Equity Research and Promotion, Philadelphia VA* [c]*Department of Global Health and Social Medicine, Harvard Medical School* [d]*Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine* [e]*Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences* [f]*Management and Development for Health* [g]*Department of Population Health, London School of Hygiene and Tropical Medicine* [h]*Department of Biostatistics, Harvard T. H. Chan School of Public Health* [i]*University of Texas School of Public Health*
[j]*Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu*

Preprint submitted to JECH *December 23, 2019*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spillover among network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64, 95% CI: 1.02,2.63) and HIV knowledge (OR=20.0, 95% CI: 3.70,125); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17, 95% CI: 1.33,3.57). The 14.2% increase in NMs' HIV knowledge was largely due to CAs feeling empowered to pass on prior knowledge, rather than transmitting new knowledge to their NMs.

Conclusions: Characteristics of social network members of PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, indicating that intervening in highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population typically not easily-accessible via other means, such as direct approach by the study team rather than someone known to the participant.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar intervention effects may spillover in a population.
- Although the high dropout rate of social network members meant we could examine factors leading to dropout, estimation of the outcome will be biased if those who dropped out were different from those who did not.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the YLL due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practices and reducing injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that lower adherence is known to often compromise their effectiveness [4]. However, neither biological nor behavioral methods are fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, researchers conducted trials among persons living with HIV (PLH) which led to increased HIV knowledge [7, 8]. These interventions were group discussions and teaching led by nurses, and motivational interview-

ing, respectively. Researchers also found that the intervention reduced HIV-risky behaviors concomitantly with an increase in HIV knowledge [8].

Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect (also known as an indirect or disseminated effect) is one person's exposure affecting another's outcome [10]. In the context of a social network spillover intervention, it means a change to someone's behavior who did not receive the intervention because they were socially connected to someone who did receive the intervention. This is distinct from what is sometimes called behavioral spillover, where change's in a person's behavior affects other behaviors of that same person [11]. For injection drug users, interventions have shown spillover effects of HIV prevention education, and subsequent reduced rates of risky behaviors [12]. What remains unknown, however, is whether or not a spillover effect exists for HIV knowledge during and after an intervention in other populations, particulalry sub-Saharan Africa.

New knowledge can come from a variety of sources, one of the most important of which is a person's so- cial network [13, 14]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [15, 16, 17]. This has been directly examined in participant-driven interventions, where initially-recruited participants directly educate members of their social network [18]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [19]. Work on diffusion through social networks has shown that spreading inter- vention effects beyond the initial study population can improve the cost-effectiveness of these interventions [20]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions. For instance, networks comprising many at-risk individuals who are HIV-negative may not be as receptive to change in behaviors as networks comprising a mix of those with and without HIV. Additionally, the networks of PLH are often difficult networks to ascertain [21], due to the continued stigma of HIV and AIDS in many settings [22]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an investigator-initiated intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective

[23]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects [24]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who receive the intervention themselves is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conducted a study on network members of PLH enrolled in the larger NAMWEZA intervention [25]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [26]. Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors are shared within social networks will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

3.1. Study Population

We analyzed social network data from the Agents of Change trial [27], which was a stepped-wedge randomized controlled trial that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices [28]. Here, we define CAs based on the *potential* for PLH to become so - by self-selecting into the study, PLH identify themselves as potential CAs, which we aim to foster through the intervention.

CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania. Participants were given a baseline questionnaire and were randomized to one of three waves in which to receive the intervention. At baseline, participants were also asked to recruit up to three members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. These network members (NMs) could be either HIV positive or negative, and they were given a baseline survey. The NM was only aware the CA was a participant in the intervention if the CA shared this information with them,

which many did not due to HIV-related stigma [29]. Each CA therefore formed a CAN-NM dyad with each NM they recruited, and if they recruited more than one NM, formed an egocentic network with multiple dyads.

For CAs, the intervention comprised 10 weekly structured sessions aimed at empowering PLH to become HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [30]. Within one month of each wave of the intervention, CAs were given follow-up surveys. The interventions lasted from November 2010 to January 2014.

For NMs, study staff did not offer any direction intervention; the CAs provided information directly to their NMs. Rather, their intervention status flowed from the intervention status of their CAs. Due to operational difficulties, each NM was surveyed one time during the study. We use this interview of NMs as the division between "exposed" and "unexposed,". At the time of an NM's follow-up interview, not all of their respective CAs had undergone the intervention. Therefore, the NMs were divided into "exposed" (N=381) and "unexposed" (N=329) groups based on whether their respective CA had completed their intervention at the time of the NM's first follow-up interview. In this way, we were able to assess the longitudinal spillover effect of the intervention net of temporal or geographical trends.

Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed at least one follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyad. NMs approached by a CA, but who did not participate in the study were not recorded since it was not feasible to obtain this information from study participants themselves. In sum, our sample comprises 662 CAs and 710 NMs, meaning each CA recruited 1.07 NMs on average out of a possible 3.

During this study, there was little loss-to-follow-up among the CAs (< 10%), but much higher among the NMs [31].

3.2. HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [32]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other

characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex practices to reduce the risk of transmission. The original population comprised three different groups: two groups of low-income women, and one of women and men receiving psychiatric treatment. These populations had a Chronbach's alpha of 0.78. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa among a convenience sample of 429 members of the African Methodist Episcopal church [33]. It has also been translated to Swahili, with only minor differential item functioning [34]. This indicates that the measure performs adequately in other, similar populations, relevant to this work. Here, the Cronbach's alpha was 0.71. Because the average baseline score of HIV knowledge was 0.80, with a preponderance of scores of 1.00, the main indicator of knowledge was not normally distributed, and therefore a linear regression was not ideal. We therefore summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

3.3. Demographic and contextual variables

In addition to HIV knowledge, we are interested in a set of demographic and contextual variables that may help explain some of the trends observed. In terms of demographic variables, we include age, sex, employment status, marital status, and self-identified HIV status. We also include education, which we dichotomize at 7 years or more - a level commensurate with elementary education. We selected this cutoff because it coincides with the millennium development goal (MDG) of increasing primary education completion [35].

Contextually, we included two additional variables: having a private source of water, and the number of persons sleeping in the participant's home. The first was also based on a MDG, and indicates participants with access to safe drinking water [35]. This is a proxy for the economic security of the participant. The number of persons sleeping in the participant's home is also a non-monetary indicator of their material and social resources [36]. These variables combined give a more thorough picture of the participant's economic status than employment alone.

3.4. Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in terms of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs could have more than one NM, we assessed statistical significance of homophily using univariate intercept-only Generalized Estimating Equations (GEE) with a Normal or Bernoulli distribution for continuous and dichotomous variables, respectively. We used exchangeable working correlation structures in these models. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing all their follow-up interviews, we fit a GEE to determine predictors of follow-up, clustered at the level of the CA. In this regression, we use all the variables listed above, as well as whether the CA was lost to follow-up (LTFU). In order to examine the association between these same variables and the time to follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM completed their baseline interview. The outcome was loss-to-follow-up. NMs who were lost to follow-up were censored at 3 months after their last interview, the point at which the study-defined criteria for LTFU was met [37].

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [26], we aimed to elucidate exactly what caused the HIV knowledge of NMs to increase - either CAs gaining knowledge through the intervention and passing it on, or the CAs being empowered by the intervention to pass on existing knowledge. As shown by VanderWeele et al (2015), social network spillover effects can be broken down into direct and indirect effects in the case of dyadic relationships (Figure 1) [38]. This method has since been used for novel evaluations of spillover effects [39]. Although previous studies showed that this same type of analysis cannot be done on full network data, the data in this study consisted of only paired individuals, the CAs and their NMs, so in this case the analysis does not result in biased estimates [40, 41].

One assumption of this analysis is that each pair is totally independent, which is violated here, if a CA recruited more than one NM - the multiple CA-NM dyads including the same CA would not be independent. To address this, we performed the analysis after randomly removing NMs until each CA had only a single NM. This resulted in removing 48 NMs, just 6.7% of the population.

The method parses social influence into direct and indirect effects.The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [18]. The natural direct effect (NDE) is the effect of receiving the intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. Although the intervention does not increase their HIV knowledge, it is still useful to the CAs, as it empowers them to become CAs in the first place. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [42]. The data are not publicly available due to the sensitive nature of HIV infection status.

3.5. Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, p¡0.0001), but were less likely to have at least 7 years of education (52.0% vs. 52.3%, p¡0.0001). Only 12.3% of NMs were HIV-positive, compared to *all* CAs ($p<0.0001$). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.0001
Female	380 (53.7%)	349 (53.9%)	0.956
Employed	490 (69.3%)	356 (55.0%)	<0.0001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.0001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.78
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.0001
Married	373 (52.7%)	338 (51.1%)	0.61
HIV Positive	87 (12.3%)	662 (100%)	<0.0001
Private source of water	309 (43.7%)	263 (39.7%)	0.20

Table 1: Demographic characteristics at baseline, with the results of an intercept-only GEE for differences. Specifically, for each CA-NM dyad, either a difference (for continuous variables) or concordance (for dichotomous variables) is calculated. These are then used as an outcome in an intercept-only GEE clustering on CA, and the p-value of the intercept is shown.

Logistic regression showed that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female (OR=1.64), having complete HIV knowledge (OR=20), being employed (OR=1.5), and being married (OR=1.72) were all significant predictors of increased odds of being followed-up. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (OR=0.81, 95% CI: 0.67,0.98) as did the NM living with HIV (OR=0.42, 95% CI: 0.18,0.99). CA having a private water source was significantly associated with increased odds of the NM being followed-up (OR=2.17, 95% CI: 1.33,3.57), even after controlling for the NM having a private source of water. This was the only CA-specific variable that significantly predicted an NM's follow-up.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married (HR=1.28) and the NM living with HIV (HR=0.71) significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to intervention via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [26]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.06%,2.0%), which is the effect the intervention had via the CA's HIV knowledge, and a natural direct effect (NDE) of 12.3% (95% CI:

Characteristic (N=459)	Adjusted OR (95% CI)	Adjusted HR (95% CI)
Characteristics of NMs		
Gender (Female)	1.64* (1.02,2.63)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	20* (3.70,125)	2.20 (0.97,5.01)
Employed	1.5* (1.04,2.5)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.81* (0.67,0.98)	0.92 (0.83,1.01)
Married	1.72* (1.04,2.86)	1.28* (1.01,1.64)
Living with HIV	0.42* (0.18,0.99)	0.71* (0.51,0.99)
Having a private source of water	0.86 (0.54,1.37)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.17* (1.33,3.57)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)
CA Lost to follow-up	1.06 (0.84,1.33)	1.02 (0.72,1.43)

Table 2: Results of multivariate logistic regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM was followed-up, and the continuous outcome of time-to-follow-up, respectively. "*" indicates significant at the $p < 0.05$ level.

6.1%,19.3%), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge. In other words, of the 12.9% increase in NM's having complete HIV knowledge, 12.3% occurred without a concomitant increase in their CA's HIV knowledge - their HIV knowledge increase was not mediated by the increase in HIV knowledge of their CA. This did not change when we used only one NM per CA.

Discussion

Based on results from the NAMWEZA trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the evidence for spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that maximal correct information transmitted from the study participants to members of their social networks.

We found many significant differences between CAs and the NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that by having CAs recruit from

their social network, we were able to recruit a set of social network members very different from those found in the waiting rooms in HIV treatment clinics (unsurprising, given that many NMs were not HIV-positive) [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample [43]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [44]. This means that future studies may need to tailor their interventions to work with the schedules of employed persons to ensure participation.

Although the study design potentially accessed a separate slice of the population than other methods vis-a´-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. We also note that less HIV knowledge and living with HIV predicted greater loss to follow-up among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE - i.e. knowledge spilled-over as a result of the CAs participating in NAMWEZA, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because CAs had a high average HIV knowledge score at baseline (80%), so there was less room for improvements in knowledge. Following the intervention, CAs' likelihood of complete HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was the CA becoming empowered through the intervention to pass on their existing knowledge to their NM [45].

This finding is important for future interventions: spillover effects of this intervention will likely carry over only one degree of separation from the CAs (opposed to spreading indefinitely in a snowball effect),

as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs receive. Because NMs do not actually receive the intervention, but only the contact with their CA, it is unlikely that they would feel empowered to become CAs themselves (particularly because they will not receive the *NAMWEZA* sessions), thus limiting the spread of the intervention. For CAs to increase the HIV knowledge of their NMs, they would only need to become empowered to share their information. However, for the NMs to increase the HIV knowledge of their own NMs, they would need to both increase their HIV knowledge, and become empowered to share it. This is less likely to happen than just becoming empowered, and so it is unlikely, though not impossible, that this effect would continue to spread in the population. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who form many bridging ties in the community would maximize the potential number of links by which spillover can occur [46]. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver the intervention to others, changing their own NMs into future CAs, the effects seen here could spillover continuously.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our final analysis may have biased our results. Even though the exposure was randomized, the loss to follow-up can result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Second, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Third, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways. Fourth, it

is possible that there was contamination between CAs if multiple CAs knew the same NM in what is known as partial interference [47] . However, given the large population of Dar es Salaam, and the small number of NMs recruited by each CA, we do not think this is a significant problem.

5. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Understanding how this then translates to how relationally close NMs are to their CAs, and the implications for follow-up may inform strategies for retention and may increase the impact of future HIV behavioral interventions. The mediation analysis presents a compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. Specifically, a similar setup to the study here combined with semi-structured interviews with CAs and NMs about their interactions with one another would help elucidate exactly how NMs increased their HIV knowledge (this population would likely need to comprise CAs who have disclosed their HIV sta- tus in order to prevent accidental disclosure). Future work should also examine whether increases in HIV knowledge translate to changes in behaviors putting individuals at risk of contracting HIV, particularly in Sub-Saharan Africa. The results presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

6. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

7. Contributorship

JL, MCSF, and JPO designed the analysis plan and drafted the initial paper. JPL, MCSF, KM, SK, IAL and JT conducted the underlying randomized trial, and designed it to incorporate social networks data and spillover information; these same authors provided feedback on the proposed analysis plan. JL and YL conducted analyses for the paper. All authors helped revise the drafts of the paper and approved the final manuscript.

8. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available.

9. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford Jr JM, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International journal of epidemiology. 2017;47(1):332–347.

[11] Truelove HB, Carrico AR, Weber EU, Raimi KT, Vandenbergh MP. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change. 2014;29:127–138.

[12] Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social network strategies to address HIV prevention and treatment continuum of care among at-risk and HIV-infected substance users: a systematic scoping review. AIDS and Behavior. 2017;21(4):1183–1207.

[13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[14] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[15] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[16] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[17] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[18] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[19] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[20] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[21] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[22] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[23] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[24] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[25] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[26] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[27] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[28] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[29] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[30] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[31] Siril HN, Kaaya SF, Fawzi MKS, Mtisi E, Somba M, Kilewo J, et al. CLINICAL outcomes and loss to follow-up among people living with HIV participating in the NAMWEZA intervention in Dar es Salaam, Tanzania: a prospective cohort study. AIDS research and therapy. 2017;14(1):18.

[32] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[33] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[34] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[35] Easterly W. How the millennium development goals are unfair to Africa. World development. 2009;37(1):26–35.

[36] Barbarin OA, Khomo N. Indicators of economic status and social capital in South African townships: What do they reveal about the material and social conditions in families of poor children? Childhood. 1997;4(2):193–222.

[37] Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to censor? American journal of epidemiology. 2017;187(3):623–632.

[38] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[39] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[40] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR. 2012;40(2):211–39.

[41] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

[42] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013 jun;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[43] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[44] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

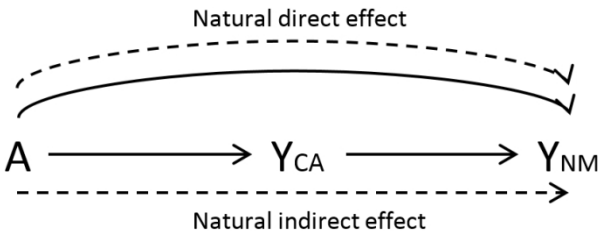
[45] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

[46] Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Social science & medicine. 2015;125:192–202.

[47] Leung MP. Treatment and spillover effects under network interference. Review of Economics and Statistics. 2016;p. 1–42.

10. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE). The NDE indicates the increase in NMs' HIV knowledge happens without a concomitant increase in their CA's HIV knowledge. The NIE, on the other hand, indicates that the increase in NMs' HIV knowledge is mediated by their CA's HIV knowledge increasing.



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759.R2
Article Type:	Original research
Date Submitted by the Author:	26-Feb-2020
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Global health, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]Perelman School of Medicine, University of Pennsylvania [b]Center for Health Equity Research and Promotion, Philadelphia VA [c]Department of Global Health and Social Medicine, Harvard Medical School [d]Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine [e]Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences [f]Management and Development for Health [g]Department of Population Health, London School of Hygiene and Tropical Medicine [h]Department of Biostatistics, Harvard T. H. Chan School of Public Health [i]University of Texas School of Public Health
[j]Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu

Preprint submitted to BMJ Open *25 February, 2020*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spillover among network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64, 95% CI: 1.02,2.63) and HIV knowledge (OR=20.0, 95% CI: 3.70,125); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17, 95% CI: 1.33,3.57). The 14.2% increase in NMs' HIV knowledge was largely due to CAs feeling empowered to pass on prior knowledge, rather than transmitting new knowledge to their NMs.

Conclusions: Characteristics of social network members of persons living with HIV PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, suggesting that intervening among highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population typically not easily-accessible via other means, such as direct approach by the study team rather than someone known to the participant.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar intervention effects may spillover in a population.
- The greater-than-ideal dropout rate of NMs was both a strength in that it allowed us to examine factors associated with dropout, but also a limitation, in that the potential of differential dropout by unmeasured factors may have biased some of our results.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the years of life lost due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practices and reducing injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that lower adherence is known to often compromise their effectiveness [4]. However, neither biological nor behavioral methods are fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals (90% diagnosis, 90% antiretroviral therapy, and 90% viral suppression among the treated) and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, researchers conducted trials among Persons Living with HIV (PLH) which led to increased HIV knowledge [7, 8]. These interventions were group discussions and teaching led by nurses, and motivational interviewing, respectively. Researchers also found that the intervention reduced HIV-risk behaviors

concomitantly with an increase in HIV knowledge [8].

Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect (also known as an indirect or disseminated effect) is one person's exposure affecting another's outcome [10]. In the context of a social network spillover intervention, it means a change to someone's behavior who did not receive the intervention because they were socially connected to someone who did receive the intervention. This is distinct from what is sometimes called behavioral spillover, where change's in a person's behavior affects other behaviors of that same person [11]. For injection drug users, HIV prevention educational interventions were demonstrated to have spillover effects of HIV prevention education, and subsequent reduced rates of risky behaviors [12]. What remains unknown, however, is whether or not a spillover effect exists for HIV knowledge during and after an intervention in other populations, particularly sub-Saharan Africa. In other words, we aim to determine that if an intervention increases someone's HIV knowledge, how members of their social networks also increase their HIV knowledge.

New knowledge can come from a variety of sources, one of the most important of which is a person's social network [13, 14]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [15, 16, 17]. This has been examined in participant-driven interventions, where initially-recruited participants educate members of their social network one-on-one [18]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [19]. Work on diffusion through social networks, how a belief or behavior can be "contagious" within a network, has shown that spreading intervention effects beyond the initial study population can improve the cost-effectiveness of these interventions [20]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions. For instance, networks comprising many at-risk individuals who are HIV-negative may not be as receptive to change in behaviors as networks comprising a mix of those with and without HIV. Additionally, the networks of PLH are often difficult networks to ascertain [21], due to the continued stigma of HIV and AIDS in many settings [22]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an investigator-initiated intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective (e.g. Respondent Driven

Sampling) [23]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects in these settings (out of approximately 750) [24]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who receive the intervention themselves is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conducted a study on network members of PLH enrolled in a behavior change intervention [25]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [26]. Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors are shared within social networks will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

3.1. Study Population

We analyzed social network data from the Agents of Change trial [27], which was a stepped-wedge randomized controlled trial [28] that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices. Here, we define CAs based on the *potential* for PLH to become so - by self-selecting into the study, PLH identify themselves as potential CAs, which we aim to foster through the intervention.

CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania, and we received written consent from each CA. Participants were given a baseline questionnaire and were randomized to one of three waves in which to receive the intervention. At baseline, participants were also asked to recruit up to three members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. We obtained written consent from these nominated network members (NMs). NMs could be either HIV positive or negative, and they were given a baseline survey. The NM was only aware the CA was a participant in the intervention if the CA shared this information with them, which

many did not due to HIV-related stigma [29]. Each CA therefore formed a CA-NM dyad with each NM they recruited, and if they recruited more than one NM, formed a set of CA-NM dyads with a common CA.

For CAs, the intervention comprised 10 weekly structured sessions aimed at empowering PLH to become HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [30]. Within one month of each wave of the intervention, CAs were given follow-up surveys. Across all waves, the interventions lasted from November 2010 to January 2014, and the final interviews were conducted in March 2014.

The NMs did not receive any intervention at any point during the study. Rather, their intervention status flowed from the intervention status of their CAs. Due to operational difficulties, each NM was surveyed two times during the study: baseline and after the first wave, rather than baseline and one after each wave. We use this interview of NMs as the division between "exposed" and "unexposed". At the time of an NM's follow-up interview, their respective CA may or may not have undergone the intervention. Therefore, the NMs were divided into "exposed" (N=381) and "unexposed" (N=329) groups based on whether their respective CA had completed their intervention at the time of the NM's follow-up interview. In this way, we were able to assess the spillover effect of the intervention net of temporal or geographical trends.

Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed at least one follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyad. NMs approached by a CA, but who did not participate in the study were not recorded since it was not feasible to obtain this information from study participants themselves. As we lost some CAs and NMs to follow up, we completed our analyses without their data, assuming it to be Missing Completely at Random (MCAR). In sum, our sample comprises 662 CAs and 710 NMs, meaning each CA recruited 1.07 NMs on average out of a possible 3, and 44 CAs nominated at least two NMs.

During this study, there was little loss-to-follow-up among the CAs (< 10%), but much higher among the NMs (36.8%) [31].

3.2. HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [32]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex

practices to reduce the risk of transmission. The original population comprised three different groups: two groups of low-income women, and one of women and men receiving psychiatric treatment. In these populations, the measure had a Chronbach's alpha of 0.78. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa among a convenience sample of 429 members of the African Methodist Episcopal church [33]. It has also been translated to Swahili, with only minor differential item functioning [34]. This indicates that the measure performs adequately in other, similar populations, relevant to this work. Here, the Cronbach's alpha was 0.71. Because the average baseline score of HIV knowledge was 0.80, with a preponderance of scores of 1.00, the main indicator of knowledge was not normally distributed, and therefore a continuous predictor was not ideal [35,36]. We therefore summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

3.3. Demographic and contextual variables

In addition to HIV knowledge, we are interested in a set of demographic and contextual variables that may help explain some of the trends observed. In terms of demographic variables, we include age, sex, employment status, marital status, and self-identified HIV status. We also include education, which we dichotomize at 7 years or more - a level commensurate with elementary education. We selected this cutoff because it coincides with the millennium development goal (MDG) of increasing primary education completion [37].

Contextually, we included two additional variables: having a private source of water, and the number of persons sleeping in the participant's home. The first was also based on a MDG, and indicates participants with access to safe drinking water [37]. This is a proxy for the economic security of the participant. The number of persons sleeping in the participant's home is also a non-monetary indicator of their material and social resources [38]. These variables combined give a more thorough picture of the participant's economic status than employment alone.

3.4. Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in terms of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs and NMs self-selected into the study and were not randomized to CA/NM status, we do not *a priori*

expect them to be completely similar. In addition, multiple NMs could share a CA and would therefore not be independent due to the shared variation and latent characteristics of having the same CA. We therefore assessed statistical significance of homophily on the set of CA-NM dyads using a permutation test. For continuous variables, the difference between the CA and NM was calculated, and for categorical variables, whether the CA and NM were concordant or discordant was recorded. We then randomly permuted CA-NM ties (keeping number of ties per CA constant), and then recalculated these statistics 1,000 times. We then examined the percentile of the observed difference relative to the permuted differences [39]. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing their follow-up interview, we fit a log-binomial regression to determine predictors of follow-up. In this regression, we use all the variables listed above, as well as whether the CA remained in the study for its full duration. In order to examine the association between these same variables and the time of follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM completed their baseline interview. The outcome here was whether the NM completed a follow-up interview. NMs who were lost to follow-up were censored at the time of their latest interview [40]. For this analysis we report Hazard Ratios (HRs), which have important limitations: selection bias and sensitivity to study period [41]. Either of these could affect this analysis, hence our use of logistic regression as a primary analysis. However, they remain useful as a sensitivity analysis.

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [27], we aimed to elucidate exactly what caused the HIV knowledge of NMs to increase - either CAs gaining knowledge through the intervention and passing it on, or the CAs being empowered by the intervention to pass on existing knowledge. As the wedge in which the CA received the *NAMWEZA* intervention was randomized, we treat the time at which an NM was potentially exposed to *NAMWEZA* through their CA as similarly randomized. This randomization scheme allowed us to explore the spillover effect of CAs' HIV knowledge onto their respective NMs via a mediation analysis.

As shown by VanderWeele et al (2015), social network spillover effects in the case of dyadic relationships can be broken down into concepts from mediation analysis: direct and indirect effects (Figure 1) [42]. This method has since been used for novel evaluations of spillover effects [43]. Although previous studies showed that this same type of analysis cannot be done on full network data, the data in this study consisted of only dyads, the CAs and their NMs, so in this case the method is appropriate [44, 45].

One assumption of this analysis is that the dyads are independent, which is violated here; if a CA recruited more than one NM, the multiple CA-NM dyads involving the same CA would not be independent. To address this, we performed the analysis after randomly removing NMs until each CA had only a single NM. This resulted in removing 48 NMs, just 6.7% of the population. We found that the point estimates were nearly identical, but that the confidence intervals were slightly larger due to the reduced sample size. No coefficients changed from significant to non-significant in this analysis (data not shown).

A second, related assumption is that of *partial interference*, that the effects in one cluster does not affect another cluster - here, one CA-NM dyad affecting another [46]. This could occur if two NMs of different CAs happen to know one-another outside of the study, one has a CA who was randomized to an earlier wedge, and shares what they know of it with the otherwise-unexposed NM. However, due to the size of Dar es Salaam, and the number of HIV treatment clinics in which recruitment occurred, we expect few CAs or NMs to know one another outside of the study, limiting the potential for partial interference.

The method parses social influence into direct and indirect effects. The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [18]. The natural direct effect (NDE) is the effect of receiving the intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. Although the intervention does not increase their HIV knowledge, it is still useful to the CAs, as it empowers them to become CAs in the first place. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [47,48]. The data are not publicly available due to the sensitive nature of HIV infection status.

3.5. Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, p<0.001), but were less likely to have at least 7 years of education (52.0% vs. 52.3%, p<0.001). Only 12.3% of NMs were HIV-positive, compared to *all* CAs (p<0.001). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.001
Female	380 (53.7%)	349 (53.9%)	0.89
Employed	490 (69.3%)	356 (55.0%)	<0.001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.65
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.001
Married	373 (52.7%)	338 (51.1%)	0.56
HIV Positive	87 (12.3%)	662 (100%)	<0.001
Private source of water	309 (43.7%)	263 (39.7%)	0.19

Table 1: Demographic characteristics at baseline, with the results of a permutation test for homophily. Specifically, for each CA- NM dyad, either a difference (for continuous variables) or concordance (for dichotomous variables) is calculated. For example, if a CA was 39 years old, and their NM was 25 years old, the difference would be 14 years old. If a CA was male and their NM was Female, the pair would be discordant for sex. CA-NM pairs were then randomly reshuffled, the edge-wise characteristics recalculated, and the observed difference compared to the distribution of randomized differences.

Logistic regression showed that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female (RR=1.44, 95% CI: 1.05,1.97), having complete HIV knowledge (RR=10, 95% CI: 2.33,42), being employed (RR=1.43, 95% CI: 1.08,1.89), and being married (RR=1.55, 95% CI: 1.03,2.33) were all significant predictors of increased odds of completing a follow-up interview. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (RR=0.85, 95% CI: 0.74,0.98) as did the NM living with HIV (RR=0.40, 95% CI: 0.17,0.96). CA having a private water source was significantly associated with increased odds of the NM being followed-up (RR=2.07, 95% CI: 1.25,3.42), even after controlling for the NM having a private source

of water. This was the only CA-specific variable that significantly predicted an NM's follow-up.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married (HR=1.28, 95% CI: 1.01,1.64) and the NM living with HIV (HR=0.71, 95% CI: 0.51,0.99) significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to intervention via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [26]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.06%,2.0%), which is the effect the intervention had via the CA's HIV knowledge, and a natural direct effect (NDE) of 12.3% (95% CI:

Characteristic (N=459)	**Adjusted RR (95% CI)**	**Adjusted HR (95% CI)**
Characteristics of NMs		
Gender (Female)	1.44* (1.05,1.97)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	10* (2.33,42)	2.20 (0.97,5.01)
Employed	1.43* (1.08,1.89)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.85* (0.74,0.98)	0.92 (0.83,1.01)
Married	1.55* (1.03,2.33)	1.28* (1.01,1.64)
Living with HIV	0.40* (0.17,0.96)	0.71* (0.51,0.99)
Having a private source of water	0.89 (0.59,1.34)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.07* (1.25,3.42)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)
CA Lost to follow-up	1.06 (0.84,1.33)	1.02 (0.72,1.43)

Table 2: Results of multivariate log-binomial regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM completed a follow-up questionnaire, and the continuous outcome of time-to-completion of follow-up questionnaire, respectively. "*" indicates significant at the $p < 0.05$ level.

6.1%,19.3%), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge. In other words, of the 12.9% increase in NM's having complete HIV knowledge, 12.3% occurred without a concomitant increase in their CA's HIV knowledge. In other words, their HIV knowledge increase was not mediated by the increase in HIV knowledge of their CA. This did not change when we used only one NM per CA.

5. Discussion

Based on results from the NAMWEZA trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the evidence for spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that intervention information is transmitted from the study participants to members of their social networks.

We found several significant differences between CAs and NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that by having CAs recruit from their social network, we were able to recruit a set of social network members different in multiple ways from those found in the waiting rooms in HIV treatment clinics (unsurprising, given that many NMs were not HIV-positive) [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample, as our sample of NMs was not composed of only those who were HIV-positive, but also many who were HIV-negative [49]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [50]. This means that future studies may need to tailor their interventions to work with the schedules of employed persons to increase participation.

Although the study design potentially accessed a separate slice of the population than other methods vis-a´-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. The one CA characteristic which did predict NMs completing a follow-up interview, having a private source of water, was a proxy for CA's socioeconomic status. This may have been because CAs with greater resources may have had more time available to pass information to their NMs, retaining the NM's interest longer [51]. We also note that less HIV knowledge and living with HIV predicted reduced likelihood of completing a follow-up questionnaire among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not

mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE; i.e. knowledge spilled-over as a result of the CAs participating in NAMWEZA, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because CAs had a high average HIV knowledge score at baseline (80%), so there was less room for improvements in knowledge. Following the intervention, CAs' likelihood of complete HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was the CA becoming empowered through the intervention to pass on their existing knowledge to their NM [52].

This finding is important for future interventions: spillover effects of this intervention will likely carry over only to those directly-connected to the CAs (opposed to spreading indefinitely in a snowball effect), as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs receive. Because NMs do not actually receive the intervention, it is unlikely that they would feel empowered to become CAs themselves (particularly because they will not receive the *NAMWEZA* sessions), thus limiting the spread of the intervention. For CAs to increase the HIV knowledge of their NMs, they would only need to become empowered to share their information. However, for the NMs to increase the HIV knowledge of their own NMs, they would need to both increase their HIV knowledge, and become empowered to share it. This is less likely to happen than just becoming empowered, and so it is unlikely, though not impossible, that this effect would continue to spread in the population. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who form many bridging ties in the community would maximize the potential number of links by which spillover can occur [53]. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver the intervention to others, changing their own NMs into future CAs, those new NMs-turned-CAs could then deliver the intervention to a second set of NMs, again empowering them to become CAs.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our final analysis may have biased our results. Even though the exposure was randomized, the loss to follow-up can result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention

because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Adjusting for censoring weights may ameliorate this issue [54]. Second, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Third, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways.

6. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Coupling this with a deeper understanding of the mental heuristics CAs used to select NMs (e.g. did CAs mentally search their close or peripheral network for those at-risk of HIV), may lead to different strategies for recruitment and retention, leading to stronger effects of behavioral interventions. The mediation analysis presents a compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. Specifically, a similar setup to the study here combined with semi-structured interviews with CAs and NMs about their interactions with one another would help elucidate exactly how NMs increased their HIV knowledge (this population would likely need to comprise CAs who have disclosed their HIV status in order to prevent accidental

disclosure). Future work should also examine whether increases in HIV knowledge translate to changes in behaviors which may increase one's risk of contracting HIV, particularly in Sub-Saharan Africa. The results presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

7. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

8. Contributorship

JL, MCSF, and JPO designed the analysis plan and drafted the initial paper. JPL, MCSF, KM, SK, IAL and JT conducted the underlying randomized trial, and designed it to incorporate social networks data and spillover information; these same authors provided feedback on the proposed analysis plan. JL and YL conducted analyses for the paper. All authors helped revise the drafts of the paper and approved the final manuscript.

9. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available. As the data contain enough information to potentially-uniquely identify specific participants, in the sensitive context of HIV, we have chosen not to make the data available.

10. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban

women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford Jr JM, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International journal of epidemiology. 2017;47(1):332–347.

[11] Truelove HB, Carrico AR, Weber EU, Raimi KT, Vandenbergh MP. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change. 2014;29:127–138.

[12] Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social network strategies to address HIV prevention and treatment continuum of care among at-risk and HIV-infected substance users: a systematic scoping review. AIDS and Behavior. 2017;21(4):1183–1207.

[13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[14] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[15] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[16] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[17] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[18] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[19] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[20] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[21] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[22] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[23] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[24] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[25] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[26] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[27] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[28] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[29] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[30] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[31] Siril HN, Kaaya SF, Fawzi MKS, Mtisi E, Somba M, Kilewo J, et al. CLINICAL outcomes and loss to follow-up among people living with HIV participating in the NAMWEZA intervention in Dar es Salaam, Tanzania: a prospective cohort study. AIDS research and therapy. 2017;14(1):18.

[32] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[33] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[34] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[35] Peltzer K, Matseke G, Mzolo T, Majaja M. Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. BMC public health. 2009;9(1):174.

[36] Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry. 2009;8(1):50-61.

[37] Easterly W. How the millennium development goals are unfair to Africa. World development. 2009;37(1):26–35.

[38] Barbarin OA, Khomo N. Indicators of economic status and social capital in South African townships: What do they reveal about the material and social conditions in families of poor children? Childhood. 1997;4(2):193–222.

[39] Smith JA, McPherson M, Smith-Lovin L. Social distance in the United States: Sex, race, religion, age,

and education homophily among confidants, 1985 to 2004. American Sociological Review. 2014 Jun;79(3):432-56.

[40] Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to censor? American journal of epidemiology. 2017;187(3):623–632.

[41] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.). 2010;21(1):13.

[42] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[43] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[44] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR. 2012;40(2):211–39.

[45] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

[46] Leung MP. Treatment and spillover effects under network interference. Review of Economics and Statistics. 2016;p. 1–42.

[47] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[48] Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013;18(2):137.

[49] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal

of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[50] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

[51] Berkman LF. Social support, social networks, social cohesion and health. Social work in health care. 2000 Sep 26;31(2):3-14.

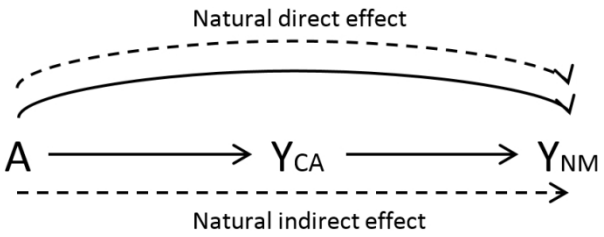
[52] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

[53] Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Social science & medicine. 2015;125:192–202.

[54] He P, Eriksson F, Scheike TH, Zhang MJ. A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. Scandinavian Journal of Statistics. 2016 Mar;43(1):103-22.

11. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE). The NDE indicates the increase in NMs' HIV knowledge happens without a concomitant increase in their CA's HIV knowledge. The NIE, on the other hand, indicates that the increase in NMs' HIV knowledge is mediated by their CA's HIV knowledge increasing.



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759.R3
Article Type:	Original research
Date Submitted by the Author:	12-Jun-2020
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Global health, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]*Perelman School of Medicine, University of Pennsylvania* [b]*Center for Health Equity Research and Promotion, Philadelphia VA* [c]*Department of Global Health and Social Medicine, Harvard Medical School* [d]*Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine* [e]*Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences* [f]*Management and Development for Health* [g]*Department of Population Health, London School of Hygiene and Tropical Medicine* [h]*Department of Biostatistics, Harvard T. H. Chan School of Public Health* [i]*University of Texas School of Public Health*
[j]*Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu*

Preprint submitted to BMJ Open *11 June, 2020*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spillover among network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64, 95% CI: 1.02,2.63) and HIV knowledge (OR=20.0, 95% CI: 3.70,125); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17, 95% CI: 1.33,3.57). The 14.2% increase in NMs' HIV knowledge was largely due to CAs feeling empowered to pass on prior knowledge, rather than transmitting new knowledge to their NMs.

Conclusions: Characteristics of social network members of persons living with HIV PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, suggesting that intervening among highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population typically not easily-accessible via other means, such as direct approach by the study team rather than someone known to the participant.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar intervention effects may spillover in a population.
- The greater-than-ideal dropout rate of NMs was both a strength in that it allowed us to examine factors associated with dropout, but also a limitation, in that the potential of differential dropout by unmeasured factors may have biased some of our results.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the years of life lost due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practices and reducing injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that lower adherence is known to often compromise their effectiveness [4]. However, neither biological nor behavioral methods are fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals (90% diagnosis, 90% antiretroviral therapy, and 90% viral suppression among the treated) and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, researchers conducted trials among Persons Living with HIV (PLH) which led to increased HIV knowledge [7, 8]. These interventions were group discussions and teaching led by nurses, and motivational interviewing, respectively. Researchers also found that the intervention reduced HIV-risk behaviors

concomitantly with an increase in HIV knowledge [8].

Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect (also known as an indirect or disseminated effect) is one person's exposure affecting another's outcome [10]. In the context of a social network spillover intervention, it means a change to someone's behavior who did not receive the intervention because they were socially connected to someone who did receive the intervention. This is distinct from what is sometimes called behavioral spillover, where change's in a person's behavior affects other behaviors of that same person [11]. For injection drug users, HIV prevention educational interventions were demonstrated to have spillover effects of HIV prevention education, and subsequent reduced rates of risky behaviors [12]. What remains unknown, however, is whether or not a spillover effect exists for HIV knowledge during and after an intervention in other populations, particularly sub-Saharan Africa. In other words, we aim to determine that if an intervention increases someone's HIV knowledge, how members of their social networks also increase their HIV knowledge.

New knowledge can come from a variety of sources, one of the most important of which is a person's social network [13, 14]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [15, 16, 17]. This has been examined in participant-driven interventions, where initially-recruited participants educate members of their social network one-on-one [18]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [19]. Work on diffusion through social networks, how a belief or behavior can be "contagious" within a network, has shown that spreading intervention effects beyond the initial study population can improve the cost-effectiveness of these interventions [20]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions. For instance, networks comprising many at-risk individuals who are HIV-negative may not be as receptive to change in behaviors as networks comprising a mix of those with and without HIV. Additionally, the networks of PLH are often difficult networks to ascertain [21], due to the continued stigma of HIV and AIDS in many settings [22]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an investigator-initiated intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective (e.g. Respondent Driven

Sampling) [23]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects in these settings (out of approximately 750) [24]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who receive the intervention themselves is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conducted a study on network members of PLH enrolled in a behavior change intervention [25]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [26]. Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors are shared within social networks will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

3.1. Study Population

We analyzed social network data from the Agents of Change trial [27], which was a stepped-wedge randomized controlled trial [28] that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices. Here, we define CAs based on the *potential* for PLH to become so - by self-selecting into the study, PLH identify themselves as potential CAs. The NAMWEZA intervention is then designed to foster a CA's ability to truly act as a Change Agent, rather than in name only.

CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania, and we received written consent from each CA. Participants were given a baseline questionnaire and were randomized to one of three waves in which to receive the intervention. At baseline, participants were also asked to recruit up to three members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. We obtained written consent from these nominated network members (NMs). NMs could be either HIV positive or negative, and they were given a baseline survey. The NM was only

aware the CA was a participant in the intervention if the CA shared this information with them, which many did not due to HIV-related stigma [29]. Each CA therefore formed a CA-NM dyad with each NM they recruited, and if they recruited more than one NM, formed a set of CA-NM dyads with a common CA.

As fits a stepped-wedge RCT, all CAs eventually received the intervention, but were randomized to *when* they received it. These waves each lasted 12 weeks, at which point the next wave began and another group of CAs received the intervention. Within each wave, the intervention comprised 10 weekly structured sessions aimed at empowering PLH to be- come HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [30]. Within one month of each wave of the intervention, CAs were given follow-up surveys. Across all waves, the interventions lasted from November 2010 to January 2014, and the final interviews were conducted in March 2014. For more information on the study design, we direct interested readers to Smith-Fawzi et al., 2019 [26].

The NMs did not receive any intervention at any point during the study. Rather, their intervention status flowed from the intervention status of their CAs. Due to operational difficulties, each NM was surveyed two times during the study: baseline and after the first wave, rather than baseline and one after each wave. In this way, all demographic and contextual variables were measured at baseline. We use this interview of NMs as the division between "exposed" and "unexposed". At the time of an NM's follow-up interview, their respective CA may or may not have undergone the intervention. Therefore, the NMs were divided into "exposed" (N=381) and "unexposed" (N=329) groups based on whether their respective CA had completed their intervention at the time of the NM's follow-up interview.

Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed at least one follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyad. NMs approached by a CA, but who did not participate in the study were not recorded since it was not feasible to obtain this information from study participants themselves. As we lost some CAs and NMs to follow up, we completed our analyses without their data, assuming it to be Missing Completely at Random (MCAR). During this study, there was little loss-to-follow-up among the CAs (< 10%), but much higher loss among the NMs (36.8%) [31]. Given an NM or CA was not lost to follow-up, complete information was available on all additional variables, including exposure, outcome, and covariates. In sum, our sample comprises 662 CAs and 710 NMs, meaning each CA recruited 1.07 NMs on average out of a possible 3, and 44 CAs nominated at least two NMs.

3.2. HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [32]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex practices to reduce the risk of transmission. The original population comprised three different groups: two groups of low-income women, and one of women and men receiving psychiatric treatment. In these populations, the measure had a Chronbach's alpha of 0.78. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa among a convenience sample of 429 members of the African Methodist Episcopal church [33]. It has also been translated to Swahili, with only minor differential item functioning [34]. This indicates that the measure performs adequately in other, similar populations, relevant to this work. Here, the Cronbach's alpha was 0.71. Because the average baseline score of HIV knowledge was 0.80, with a preponderance of scores of 1.00, the main indicator of knowledge was not normally distributed, and therefore a continuous predictor was not ideal [35,36]. We therefore summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

3.3. Demographic and contextual variables

In addition to HIV knowledge, we are interested in a set of demographic and contextual variables that may help explain some of the trends observed. In terms of demographic variables, we include age, sex, employment status, marital status, and self-identified HIV status. We also include education, which we dichotomize at 7 years or more - a level commensurate with elementary education. We selected this cutoff because it coincides with the millennium development goal (MDG) of increasing primary education completion [37].

Contextually, we included two additional variables: having a private source of water, and the number of persons sleeping in the participant's home. The first was also based on a MDG, and indicates participants with access to safe drinking water [37]. This is a proxy for the economic security of the participant. The number of persons sleeping in the participant's home is also a non-monetary indicator of their material and social resources [38]. These variables combined give a more thorough picture of the participant's economic status than employment alone.

3.4. Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in terms of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs and NMs self-selected into their respective group, and randomization only occurred *within groups*, differences between the groups were to be expected. However, we only examined homophily of baseline characteristics rather than of outcome, because comparison of outcomes between CAs and NMs would remove the benefits of randomization. Because CAs and NMs self-selected into the study and were not randomized to CA/NM status, we do not *a priori* expect them to be completely similar. In addition, multiple NMs could share a CA and would therefore not be independent due to the shared variation and latent characteristics of having the same CA. We therefore assessed statistical significance of homophily on the set of CA-NM dyads using a permutation test, a non-parametric test which builds a distribution directly from the data. For continuous variables, the difference between the CA and NM was calculated, and for categorical variables, whether the CA and NM were concordant or discordant was recorded. We then randomly permuted CA-NM ties (keeping number of ties per CA constant), and then recalculated the difference or percent concordant, respectively, 1,000 times. We then examined the percentile of the observed difference relative to the permuted differences [39]. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing their follow-up interview, we fit a log-binomial regression to determine predictors of follow-up. In this regression, we use all the variables listed above, as well as whether the CA remained in the study for its full duration. In order to examine the association between these same variables and the time of follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM completed their baseline interview. The outcome here was whether the NM completed a follow-up interview. NMs who were lost to follow-up were censored at the time of their latest interview [40,41].

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [27], we aimed to elucidate exactly what caused the HIV knowledge of NMs to increase - either CAs gaining knowledge through the intervention and passing it on, or the CAs being empowered by the intervention to pass on existing knowledge. As the wedge in which the CA received the *NAMWEZA* intervention was randomized, we treat the time at which an NM was potentially exposed to *NAMWEZA* through their CA as similarly randomized. This randomization scheme allowed us to explore the spillover effect of CAs' HIV knowledge onto their respective NMs via a mediation analysis. These pathways represent

different types of spillover effects: the exposure or outcome of one person affecting the outcome of another person.

As shown by VanderWeele et al (2015), social network spillover effects in the case of dyadic relationships can be broken down into concepts from mediation analysis: direct and indirect effects (Figure 1) [42]. This method has since been used for novel evaluations of spillover effects [43,44]. Although previous studies showed that this same type of analysis cannot be done on full network data, the data in this study consisted of only dyads, the CAs and their NMs, so in this case the method is appropriate [45, 46].

The method parses social influence into direct and indirect effects. The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [18]. The natural direct effect (NDE) is the effect of receiving the intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. Although the intervention does not increase their HIV knowledge, it is still useful to the CAs, as it empowers them to become CAs in the first place. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [47,48].

One assumption of this analysis is that the dyads are independent, which is violated here; if a CA recruited more than one NM, the multiple CA-NM dyads involving the same CA would not be independent. To address this, we performed the analysis after randomly removing NMs until each CA had only a single NM. This resulted in removing 48 NMs, just 6.7% of the population. We found that the point estimates were nearly identical, but that the confidence intervals were slightly larger due to the reduced sample size. No coefficients changed from significant to non-significant in this analysis (data not shown).

A second, related assumption is that of *partial interference*, that the effects in one cluster does not affect another cluster - here, one CA-NM dyad affecting another [49]. This could occur if two NMs of different CAs happen to know one-another outside of the study, one has a CA who was randomized to an earlier wedge, and shares what they know of it with the otherwise-unexposed NM. However, due to the large size of Dar es Salaam, and the number of HIV treatment clinics in which recruitment occurred, we expect few CAs or NMs to know one another outside of the study (other than NMs knowing the CA who recruited

them), limiting the potential for partial interference.

The data are not publicly available due to the sensitive nature of HIV infection status.

3.5. Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, p<0.001), but were less likely to have at least 7 years of education (52.0% vs. 52.3%, p<0.001). Only 12.3% of NMs were HIV-positive, compared to *all* CAs (p<0.001). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.001
Female	380 (53.7%)	349 (53.9%)	0.89
Employed	490 (69.3%)	356 (55.0%)	<0.001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.65
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.001
Married	373 (52.7%)	338 (51.1%)	0.56
HIV Positive	87 (12.3%)	662 (100%)	N/A
Private source of water	309 (43.7%)	263 (39.7%)	0.19

Table 1: Demographic characteristics at baseline, with the results of a permutation test for homophily. Specifically, for each CA- NM dyad, either a difference (for continuous variables) or concordance (for dichotomous variables) is calculated. For example, if a CA was 39 years old, and their NM was 25 years old, the difference would be 14 years old. If a CA was male and their NM was Female, the pair would be discordant for sex. CA-NM pairs were then randomly reshuffled, the edge-wise characteristics recalculated, and the observed difference compared to the distribution of randomized differences.

Logistic regression showed that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female (RR=1.44, 95% CI: 1.05,1.97), having complete HIV knowledge (RR=10, 95% CI: 2.33,42), being employed (RR=1.43, 95% CI: 1.08,1.89), and being married (RR=1.55, 95% CI: 1.03,2.33) were all significant predictors of increased odds of completing a follow-up interview. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (RR=0.85, 95% CI: 0.74,0.98) as did the NM living with HIV (RR=0.40, 95% CI: 0.17,0.96). CA having a private water source was significantly associated with increased odds of the NM being followed-up (RR=2.07, 95% CI: 1.25,3.42), even after controlling for the NM having a private source of water. This was the only CA-specific variable that significantly predicted an NM's follow-up.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married (HR=1.28, 95% CI: 1.01,1.64) and the NM living with HIV (HR=0.71, 95% CI: 0.51,0.99) significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to intervention via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [26]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.06%,2.0%), which is the effect the intervention had via the CA's HIV knowledge, and a natural direct effect (NDE) of 12.3% (95% CI: 6.1%,19.3%), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge. In other words, of the 12.9% increase in NM's having complete HIV knowledge, 12.3% occurred without a concomitant increase in their CA's HIV knowledge. In other words, their HIV knowledge increase was not mediated by the increase in HIV knowledge of their CA. This did not change when we used only one NM per CA.

5. Discussion

Based on results from the NAMWEZA trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the evidence for spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that intervention information is transmitted from the study participants to members of their social networks.

Characteristic (N=459)	Adjusted RR (95% CI)	Adjusted HR (95% CI)
Characteristics of NMs		
Gender (Female)	1.44* (1.05,1.97)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	10* (2.33,42)	2.20 (0.97,5.01)
Employed	1.43* (1.08,1.89)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.85* (0.74,0.98)	0.92 (0.83,1.01)
Married	1.55* (1.03,2.33)	1.28* (1.01,1.64)
Living with HIV	0.40* (0.17,0.96)	0.71* (0.51,0.99)
Having a private source of water	0.89 (0.59,1.34)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.07* (1.25,3.42)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)
CA Lost to follow-up	1.06 (0.84,1.33)	1.02 (0.72,1.43)

Table 2: Results of multivariate log-binomial regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM completed a follow-up questionnaire, and the continuous outcome of time-to-completion of follow-up questionnaire, respectively. "*" indicates significant at the $p < 0.05$ level.

We found several significant differences between CAs and NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that by having CAs recruit from their social network, we were able to recruit a set of social network members different in multiple ways from those found in the waiting rooms in HIV treatment clinics (unsurprising, given that many NMs were not HIV-positive) [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample, as our sample of NMs was not composed of only those who were HIV-positive, but also many who were HIV-negative [50]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [51]. This means that future studies may need to tailor their interventions to work with the schedules of employed persons to increase participation.

Although the study design potentially accessed a separate slice of the population than other methods vis-a´-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA

characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. The one CA characteristic which did predict NMs completing a follow-up interview, having a private source of water, was a proxy for CA's socioeconomic status. This may have been because CAs with greater resources may have had more time available to pass information to their NMs, retaining the NM's interest longer [52]. We also note that less HIV knowledge and living with HIV predicted reduced likelihood of completing a follow-up questionnaire among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE; i.e. knowledge spilled-over as a result of the CAs participating in NAMWEZA, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because CAs had a high average HIV knowledge score at baseline (80%), so there was less room for improvements in knowledge. Following the intervention, CAs' likelihood of complete HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was the CA becoming empowered through the intervention to pass on their existing knowledge to their NM [53].

This finding is important for future interventions: spillover effects of this intervention will likely carry over only to those directly-connected to the CAs (opposed to spreading indefinitely in a snowball effect), as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs receive. Because NMs do not actually receive the intervention, it is unlikely that they would feel empowered to become CAs themselves (particularly because they will not receive the *NAMWEZA* sessions), thus limiting the spread of the intervention. For CAs to increase the HIV knowledge of their NMs, they would only need to become empowered to share their information. However, for the NMs to increase the HIV knowledge of their own NMs, they would need to both increase their HIV knowledge, and become empowered to share it. This is less likely to happen than just becoming empowered, and so it is unlikely, though not impossible, that this effect would continue to spread in the population. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who form many bridging ties in the community would maximize the potential number of links by which spillover can occur [54]. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver

the intervention to others, changing their own NMs into future CAs, those new NMs-turned-CAs could then deliver the intervention to a second set of NMs, again empowering them to become CAs.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our analysis of spillover effects may be biased. Even though the exposure was randomized, the loss to follow-up can result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Adjusting for censoring weights may ameliorate this issue [55]. Second, our use of Hazard Ratios (HRs) has important limitations: they are subject to selection bias, are sensitive to study period, and only provide one estimate during the study [41]. Any of these limitations could affect this analysis, hence our use of logistic regression as a primary analysis. However, they remain useful as a sensitivity analysis. Third, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Fourth, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways.

6. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Coupling this with a deeper understanding of the mental heuristics CAs used to select NMs (e.g. did CAs mentally search their close or peripheral network for those at-risk of HIV), may lead to different strategies for recruitment and retention, leading to stronger effects of behavioral interventions. The mediation analysis presents a

compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. Specifically, a similar setup to the study here combined with semi-structured interviews with CAs and NMs about their interactions with one another would help elucidate exactly how NMs increased their HIV knowledge (this population would likely need to comprise CAs who have disclosed their HIV status in order to prevent accidental disclosure). Future work should also examine whether increases in HIV knowledge translate to changes in behaviors which may increase one's risk of contracting HIV, particularly in Sub-Saharan Africa. The results presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

7. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

8. Contributorship

JL, MCSF, and JPO designed the analysis plan and drafted the initial paper. JPL, MCSF, KM, SK, IAL and JT conducted the underlying randomized trial, and designed it to incorporate social networks data and spillover information; these same authors provided feedback on the proposed analysis plan. JL and YL conducted analyses for the paper. All authors helped revise the drafts of the paper and approved the final manuscript.

9. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available. As the data contain enough information to potentially-uniquely identify specific participants, in the sensitive context of HIV, we have chosen not to make the data available.

10. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford Jr JM, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International journal of epidemiology. 2017;47(1):332–347.

[11] Truelove HB, Carrico AR, Weber EU, Raimi KT, Vandenbergh MP. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change. 2014;29:127–138.

[12] Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social network strategies to address HIV prevention and treatment continuum of care among at-risk and HIV-infected substance users: a systematic scoping review. AIDS and Behavior. 2017;21(4):1183–1207.

[13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[14] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[15] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[16] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[17] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[18] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[19] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[20] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[21] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[22] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[23] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[24] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[25] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[26] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[27] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[28] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[29] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[30] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[31] Siril HN, Kaaya SF, Fawzi MKS, Mtisi E, Somba M, Kilewo J, et al. CLINICAL outcomes and loss to follow-up among people living with HIV participating in the NAMWEZA intervention in Dar es Salaam, Tanzania: a prospective cohort study. AIDS research and therapy. 2017;14(1):18.

[32] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[33] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[34] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[35] Peltzer K, Matseke G, Mzolo T, Majaja M. Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. BMC public health. 2009;9(1):174.

[36] Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharmaceutical Statistics: The

Journal of Applied Statistics in the Pharmaceutical Industry. 2009;8(1):50-61.

[37] Easterly W. How the millennium development goals are unfair to Africa. World development. 2009;37(1):26–35.

[38] Barbarin OA, Khomo N. Indicators of economic status and social capital in South African townships: What do they reveal about the material and social conditions in families of poor children? Childhood. 1997;4(2):193–222.

[39] Smith JA, McPherson M, Smith-Lovin L. Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. American Sociological Review. 2014 Jun;79(3):432-56.

[40] Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to censor? American journal of epidemiology. 2017;187(3):623–632.

[41] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.). 2010;21(1):13.

[42] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[43] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[44] Imai K, Jiang Z. Identification and sensitivity analysis of contagion effects in randomized placebo-controlled trials. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2019.

[45] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR. 2012;40(2):211–39.

[46] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

[47] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized

trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[48] Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013;18(2):137.

[49] Leung MP. Treatment and spillover effects under network interference. Review of Economics and Statistics. 2016;p. 1–42.

[50] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[51] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

[52] Berkman LF. Social support, social networks, social cohesion and health. Social work in health care. 2000 Sep 26;31(2):3-14.

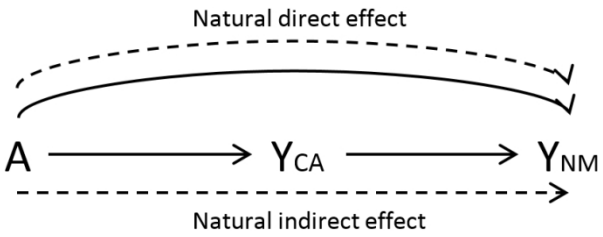
[53] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

[54] Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Social science & medicine. 2015;125:192–202.

[55] He P, Eriksson F, Scheike TH, Zhang MJ. A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. Scandinavian Journal of Statistics. 2016 Mar;43(1):103-22.

11. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE). The NDE indicates the increase in NMs' HIV knowledge happens without a concomitant increase in their CA's HIV knowledge. The NIE, on the other hand, indicates that the increase in NMs' HIV knowledge is mediated by their CA's HIV knowledge increasing. Solid lines indicate paths of causality between variables. Dashed lines represent the line or lines composing the effect of interest.



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759.R4
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2020
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Global health, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]*Perelman School of Medicine, University of Pennsylvania* [b]*Center for Health Equity Research and Promotion, Philadelphia VA* [c]*Department of Global Health and Social Medicine, Harvard Medical School* [d]*Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine* [e]*Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences* [f]*Management and Development for Health* [g]*Department of Population Health, London School of Hygiene and Tropical Medicine* [h]*Department of Biostatistics, Harvard T. H. Chan School of Public Health* [i]*University of Texas School of Public Health*
[j]*Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu*

Preprint submitted to BMJ Open *8 July, 2020*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spillover among network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64, 95% CI: 1.02,2.63) and HIV knowledge (OR=20.0, 95% CI: 3.70,125); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17, 95% CI: 1.33,3.57). The 14.2% increase in NMs' HIV knowledge was largely due to CAs feeling empowered to pass on prior knowledge, rather than transmitting new knowledge to their NMs.

Conclusions: Characteristics of social network members of persons living with HIV PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, suggesting that intervening among highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population typically not easily-accessible via other means, such as direct approach by the study team rather than someone known to the participant.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar intervention effects may spillover in a population.
- The greater-than-ideal dropout rate of NMs was both a strength in that it allowed us to examine factors associated with dropout, but also a limitation, in that the potential of differential dropout by unmeasured factors may have biased some of our results.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the years of life lost due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practices and reducing injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that lower adherence is known to often compromise their effectiveness [4]. However, neither biological nor behavioral methods are fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals (90% diagnosis, 90% antiretroviral therapy, and 90% viral suppression among the treated) and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, researchers conducted trials among Persons Living with HIV (PLH) which led to increased HIV knowledge [7, 8]. These interventions were group discussions and teaching led by nurses, and motivational

interviewing, respectively. Researchers also found that the intervention reduced HIV-risk behaviors concomitantly with an increase in HIV knowledge [8].

1. Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect (also known as an indirect or disseminated effect) is one person's exposure affecting another's outcome [10]. In the context of a social network spillover intervention, it corresponds to an individual who was unexposed to an intervention changing their behavior because they were socially-connected to an individual who did receive the intervention. This is distinct from what is sometimes called behavioral spillover, where changes in a person's behavior affect other behaviors of that same person [11]. For injection drug users, HIV prevention educational interventions were demonstrated to have spillover effects of HIV prevention education, and subsequent reduced rates of risky behaviors [12]. Studies have also used proxy variables for social network ties such as inviting social network members to watch educational programming [13] or time spent shopping at the market [14] to evaluate spillover effects for HIV knowledge, generally finding evidence for spillover. However, spillover in HIV knowledge between known social network ties generally remains understudied, particularly in sub-Saharan Africa [15]. We therefore aim to determine whether social network members those receiving an HIV behavioral/knowledge intervention also increase their HIV knowledge.

New knowledge can come from a variety of sources, one of the most important of which is a person's social network [16, 17]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [18, 19, 20]. This has been examined in participant-driven interventions, where initially-recruited participants educate members of their social network one-on-one [21]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [22]. Work on diffusion through social networks, how a belief or behavior can be "contagious" within a network, has shown that spreading intervention effects beyond the initial study population can improve the cost-effectiveness of these interventions [23]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions. For instance, networks comprising many at-risk individuals who are HIV-negative may not be as receptive to change in behaviors as networks comprising a mix of those with and without HIV. Additionally, the networks of PLH are often difficult networks to ascertain [24], due to the continued stigma of HIV and AIDS

in many settings [25]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an investigator-initiated intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective (e.g. Respondent Driven Sampling) [26]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects in these settings (out of approximately 750) [27]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who receive the intervention themselves is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conduct a study on network members of PLH enrolled in a behavior change intervention [28]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [29]. Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors are shared within social networks will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

13.1.Study Population

We analyze social network data from the Agents of Change trial [30], which was a stepped-wedge randomized controlled trial [31] that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices. Here, we define CAs based on the *potential* for PLH to become so - by self-selecting into the study, PLH identify themselves as potential CAs. Although we refer to them as 'CAs' throughout, participants in the trial enrolled with varying levels of ability to act as a Change Agent. Through receiving the *NAMWEZA* intervention, we hypothesize that CAs will be able to truly self-actualize and subsequently act as Change Agents in their community.

CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania, and we received written consent from each CA. Participants completed a baseline questionnaire and

were randomized to one of three waves in which to receive the intervention. At baseline, participants were also asked to recruit up to three members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. We obtained written consent from these nominated network members (NMs). NMs could be either HIV positive or negative, and they were given a baseline survey. The NM was only aware the CA was a participant in the intervention if the CA shared this information with them, which many did not due to HIV-related stigma [32]. Each CA therefore formed a CA-NM dyad with each NM they recruited, and if they recruited more than one NM, formed a set of CA-NM dyads with a common CA.

As fits a stepped-wedge RCT, all CAs eventually received the intervention, but were randomized to *when* they received it. These waves each lasted 12 weeks, at which point the next wave began and another group of CAs received the intervention. Within each wave, the intervention comprised 10 weekly structured sessions aimed at empowering PLH to be- come HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [33]. Within one month of each wave of the intervention, CAs were given follow-up surveys. Across all waves, the interventions lasted from November 2010 to January 2014, and the final interviews were conducted in March 2014. For more information on the study design, we direct interested readers to Smith-Fawzi et al., 2019 [29].

The NMs did not receive any intervention at any point during the study. Rather, their intervention status flowed from the intervention status of their CAs. Due to operational difficulties, each NM was surveyed two times during the study: baseline and after the first wave, rather than baseline and one after each wave. In this way, all demographic and contextual variables were measured at baseline. We use this interview of NMs as the division between "exposed" and "unexposed". At the time of an NM's follow-up interview, their respective CA may or may not have undergone the intervention. Therefore, the NMs were divided into "exposed" (N=381) and "unexposed" (N=329) groups based on whether their respective CA had completed their intervention at the time of the NM's follow-up interview.

Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed at least one follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyad. NMs approached by a CA, but who did not participate in the study were not recorded since it was not feasible to obtain this information from study participants themselves. As we lost some CAs and NMs to follow up, we completed our analyses without their data, assuming it to be Missing Completely at Random (MCAR). During this study, there was little loss-to-follow-

up among the CAs (< 10%), but much higher loss among the NMs (36.8%) [34]. Given an NM or CA was not lost to follow-up, complete information was available on all additional variables, including exposure, outcome, and covariates. In sum, our sample comprises 662 CAs and 710 NMs, meaning each CA recruited 1.07 NMs on average out of a possible 3, and 44 CAs nominated at least two NMs.

13.2.HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [35]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex practices to reduce the risk of transmission. The original population comprised three different groups: two groups of low-income women, and one of women and men receiving psychiatric treatment. In these populations, questions on the measure had a Chronbach's alpha of 0.78. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa among a convenience sample of 429 members of the African Methodist Episcopal church [36]. It has also been translated to Swahili, with only minor differential item functioning [37]. This indicates that the measure performs adequately in other, similar populations. In the present study the Cronbach's alpha is 0.71. Because the average baseline score of HIV knowledge was 0.80, with a preponderance of scores of 1.00, the main indicator of knowledge was not normally distributed, and therefore a continuous predictor was not ideal [38,39]. We therefore summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

13.3.Demographic and contextual variables

In addition to HIV knowledge, we are interested in a set of demographic and contextual variables that may help explain some of the trends observed. In terms of demographic variables, we include age, sex, employment status, marital status, and self-identified HIV status. We also include education, which we dichotomize at 7 years or more - a level commensurate with elementary education. We selected this cutoff because it coincides with the millennium development goal (MDG) of increasing primary education completion [40].

Contextually, we included two additional variables: having a private source of water, and the number of persons sleeping in the participant's home. The first was also based on a MDG, and indicates participants

with access to safe drinking water [40]. This is a proxy for the economic security of the participant. The number of persons sleeping in the participant's home is also a non-monetary indicator of their material and social resources [41]. These variables combined give a more thorough picture of the participant's economic status than employment alone.

13.4.Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in terms of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs and NMs self-selected into their respective group, and only CAs were directly randomized (with their NMs being randomized along with them), differences between the groups were to be expected. However, we only examined homophily of baseline characteristics rather than of outcome, because comparison of outcomes between CAs and NMs would remove the benefits of randomization. Because CAs and NMs self-selected into the study and were not randomized to CA/NM status, we do not *a priori* expect them to be completely similar. In addition, multiple NMs could share a CA and would therefore not be independent due to the shared variation and latent characteristics of having the same CA. We therefore assessed statistical significance of homophily on the set of CA-NM dyads using a permutation test, a non-parametric test which has no distributional assumptions. For continuous variables, the difference between the CA and NM was calculated, and for categorical variables, whether the CA and NM were concordant or discordant was recorded. We then randomly permuted CA-NM ties (keeping number of ties per CA constant), and then recalculated the difference or percent concordant, respectively, 1,000 times. We then examine the percentile of the observed difference relative to the permuted differences [42]. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing their follow-up interview, we fit a log-binomial regression to determine predictors of follow-up. In this regression, we use all the variables listed above, as well as whether the CA remained in the study for its full duration. In order to examine the association between these same variables and the time of follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM completed their baseline interview. The outcome here was whether the NM completed a follow-up interview. NMs who were lost to follow-up were censored at the time of their latest interview [43,44].

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [30], we aim to elucidate exactly what caused the HIV knowledge of NMs to increase - either CAs

gaining knowledge through the intervention and sharing it, or the CAs being empowered by the intervention to share existing knowledge. As the wedge in which the CA received the *NAMWEZA* intervention was randomized, we treat each NM as being randomized to exposure to *NAMWEZA* at the same time as their CA. This randomization scheme allows us to explore the spillover effect of CAs' HIV knowledge onto their respective NMs via a mediation analysis. These pathways represent different types of spillover effects: the exposure or outcome of one person affecting the outcome of another person.

As shown by VanderWeele et al (2015), social network spillover effects in the case of dyadic relationships can be broken down into concepts from mediation analysis: direct and indirect effects (Figure 1) [45]. This method has since been used for novel evaluations of spillover effects [46,47,48]. Although previous studies showed that this same type of analysis cannot be done on full network data, the data in this study consisted of only dyads, the CAs and their NMs, so in this case the method is appropriate [49,50].

The method parses social influence into direct and indirect effects. The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [21]. The natural direct effect (NDE) is the effect of receiving the intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. Although the intervention does not increase their HIV knowledge, it is still useful to the CAs, as it empowers them to act as CAs in their community. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [51,52].

Importantly, this analysis requires a number of assumptions and applies to dyads only when these assumptions are met. One assumption of this analysis is that the dyads are independent, which is violated here; if a CA recruited more than one NM, the multiple CA-NM dyads involving the same CA would not be independent. To address this, we performed the analysis after randomly removing NMs until each CA had only a single NM. This resulted in removing 48 NMs, just 6.7% of the population. We found that the point estimates were nearly identical, but that the confidence intervals were slightly larger due to the reduced sample size. No coefficients changed from significant to non-significant in this analysis (data not shown).

A second, related assumption is that of *partial interference*, that the effects in one cluster does not affect another cluster - here, one CA-NM dyad affecting another [53]. This could occur if two NMs of different CAs happen to know one-another outside of the study, one has a CA who was randomized to an earlier wedge, and shares what they know of it with the otherwise-unexposed NM. However, due to the large size of Dar es Salaam, and the number of HIV treatment clinics in which recruitment occurred, we expect few CAs or NMs to know one another outside of the study (other than NMs knowing the CA who recruited them), limiting the potential for partial interference.

A third assumption of this analysis is that the outcomes of the CA and NM are independent conditional on the CA's exposure, or conditional on the CA's exposure and any confounding variables [45,46]. Because CA-NM pairs self-select and are not randomized, we do not expect these outcomes to be independent conditional of the CA's exposure, and so we adjust for additional variables to meet this assumption. In our analysis, we therefore adjust for all the variables used in the log-binomial regression.

The data are not publicly available due to the sensitive nature of HIV infection status and the socially-networked nature of the data. Because the data include specific information on social ties, some of whom have not disclosed HIV sero-status to one another, the risks of individual identification and compromising HIV sero-status are greatly increased.

13.5.Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, $p<0.001$), but were less likely to have at least 7 years of education (52.0% vs. 52.3%, $p<0.001$). Only

12.3% of NMs were HIV-positive, compared to *all* CAs ($p<0.001$). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.001
Female	380 (53.7%)	349 (53.9%)	0.89
Employed	490 (69.3%)	356 (55.0%)	<0.001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.65
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.001
Married	373 (52.7%)	338 (51.1%)	0.56
HIV Positive	87 (12.3%)	662 (100%)	N/A
Private source of water	309 (43.7%)	263 (39.7%)	0.19

Table 1: Demographic characteristics at baseline, with the results of a permutation test for homophily. Specifically, for each CA- NM dyad, either a difference (for continuous variables) or concordance (for dichotomous variables) is calculated. For example, if a CA was 39 years old, and their NM was 25 years old, the difference would be 14 years old. If a CA was male and their NM was Female, the pair would be discordant for sex. CA-NM pairs were then randomly reshuffled, the edge-wise characteristics recalculated, and the observed difference compared to the distribution of randomized differences.

Risk ratios (RRs) obtained via log-binomial regression indicated that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female (RR=1.44, 95% CI: 1.05,1.97), having complete HIV knowledge (RR=10, 95% CI: 2.33,42), being employed (RR=1.43, 95% CI: 1.08,1.89), and being married (RR=1.55, 95% CI: 1.03,2.33) were all significant predictors of increased odds of completing a follow-up interview. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (RR=0.85, 95% CI: 0.74,0.98) as did the NM living with HIV (RR=0.40, 95% CI: 0.17,0.96). CA having a private water source was significantly associated with increased odds of the NM being followed-up (RR=2.07, 95% CI: 1.25,3.42), even after controlling for the NM having a private source of water. This was the only CA-specific variable that significantly predicted an NM's follow-up.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married (HR=1.28, 95% CI: 1.01,1.64) and the NM living with HIV (HR=0.71, 95% CI: 0.51,0.99) significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to intervention via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [26]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.06%,2.0%), which is the effect the intervention had via the CA's HIV knowledge, and a natural direct effect (NDE) of 12.3% (95% CI: 6.1%,19.3%), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge. In other words, of the 12.9% increase in NM's having complete HIV knowledge, 12.3% occurred without a concomitant increase in their CA's HIV knowledge. In other words, their HIV knowledge increase was not mediated by the increase in HIV knowledge of their CA. This did not change when we used only one NM per CA.

5. Discussion

Based on results from the NAMWEZA trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the evidence for spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that intervention information is transmitted from the study participants to members of their social networks.

Characteristic (N=459)	Adjusted RR (95% CI)	Adjusted HR (95% CI)
Characteristics of NMs		
Gender (Female)	1.44* (1.05,1.97)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	10* (2.33,42)	2.20 (0.97,5.01)
Employed	1.43* (1.08,1.89)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.85* (0.74,0.98)	0.92 (0.83,1.01)
Married	1.55* (1.03,2.33)	1.28* (1.01,1.64)
Living with HIV	0.40* (0.17,0.96)	0.71* (0.51,0.99)
Having a private source of water	0.89 (0.59,1.34)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.07* (1.25,3.42)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)
CA Lost to follow-up	1.06 (0.84,1.33)	1.02 (0.72,1.43)

Table 2: Results of multivariate log-binomial regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM completed a follow-up questionnaire, and the continuous outcome of time-to-completion of follow-up questionnaire, respectively. "*" indicates significant at the $p < 0.05$ level.

We found several significant differences between CAs and NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that by having CAs recruit from their social network, we were able to recruit a set of social network members different in multiple ways from those found in the waiting rooms in HIV treatment clinics (unsurprising, given that many NMs were not HIV-positive) [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample, as our sample of NMs was not composed of only those who were HIV-positive, but also many who were HIV-negative [54]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [55]. This means that future studies may need to tailor their interventions to work with the schedules of employed persons to increase participation.

Although the study design potentially accessed a separate slice of the population than other methods vis-a´-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. The one CA characteristic which did predict NMs completing a follow-up interview, having a private source of water, was a proxy for CA's socioeconomic status. This may have been because CAs with greater resources may have had more time available to pass information to their NMs, retaining the NM's interest longer [56]. We also note that less HIV knowledge and living with HIV predicted reduced likelihood of completing a follow-up questionnaire among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE; i.e. knowledge spilled-over as a result of the CAs participating in NAMWEZA, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because CAs had a high average

HIV knowledge score at baseline (80%), so there was less room for improvements in knowledge. Following the intervention, CAs' likelihood of complete HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was the CA becoming empowered through the intervention to pass on their existing knowledge to their NM [57].

This finding is important for future interventions: spillover effects of this intervention will likely carry over only to those directly-connected to the CAs (opposed to spreading indefinitely in a snowball effect), as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs receive. Because NMs do not actually receive the intervention, it is unlikely that they would feel empowered to become CAs themselves (particularly because they will not receive the *NAMWEZA* sessions), thus limiting the spread of the intervention. For CAs to increase the HIV knowledge of their NMs, they would only need to become empowered to share their information. However, for the NMs to increase the HIV knowledge of their own NMs, they would need to both increase their HIV knowledge, and become empowered to share it. This is less likely to happen than just becoming empowered, and so it is unlikely, though not impossible, that this effect would continue to spread in the population. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who form many bridging ties in the community would maximize the potential number of links by which spillover can occur [58]. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver the intervention to others, changing their own NMs into future CAs, those new NMs-turned-CAs could then deliver the intervention to a second set of NMs, again empowering them to become CAs.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our analysis of spillover effects may be biased. Even though the exposure was randomized, the loss to follow-up can result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Adjusting for censoring weights may ameliorate this issue [59]. Second, our use of Hazard Ratios (HRs) has important limitations: they are subject to selection bias, are sensitive to study period, and only provide one estimate during the study [44]. Any of these limitations could affect this analysis, hence our use of logistic regression

as a primary analysis. However, they remain useful as a sensitivity analysis. Third, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Fourth, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways.

6. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Coupling this with a deeper understanding of the mental heuristics CAs used to select NMs (e.g. did CAs mentally search their close or peripheral network for those at-risk of HIV), may lead to different strategies for recruitment and retention, leading to stronger effects of behavioral interventions. The mediation analysis presents a compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. Specifically, a similar setup to the study here combined with semi-structured interviews with CAs and NMs about their interactions with one another would help elucidate exactly how NMs increased their HIV knowledge (this population would likely need to comprise CAs who have disclosed their HIV status in order to prevent accidental disclosure). Future work should also examine whether increases in HIV knowledge translate to changes in behaviors which may increase one's risk of contracting HIV, particularly in Sub-Saharan Africa. The results

presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

7. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

8. Contributorship

JL, MCSF, and JPO designed the analysis plan and drafted the initial paper. JPL, MCSF, KM, SK, IAL and JT conducted the underlying randomized trial, and designed it to incorporate social networks data and spillover information; these same authors provided feedback on the proposed analysis plan. JL and YL conducted analyses for the paper. All authors helped revise the drafts of the paper and approved the final manuscript.

9. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available. As the data contain enough information to potentially-uniquely identify specific participants, in the sensitive context of HIV, we have chosen not to make the data available.

10. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford Jr JM, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International journal of epidemiology. 2017;47(1):332–347.

[11] Truelove HB, Carrico AR, Weber EU, Raimi KT, Vandenbergh MP. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change. 2014;29:127–138.

[12] Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social network strategies to address HIV prevention and treatment continuum of care among at-risk and HIV-infected substance users: a systematic scoping review. AIDS and Behavior. 2017;21(4):1183–1207.

[13] Benjamin-Chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, Tran D, Arnold BF, Hubbard AE, Luby SP, and Miguel E, 2017. Spillover effects on health outcomes in low-and middle-income countries: a systematic review. International journal of epidemiology, 46(4), pp.1251-1276.

[14] Self S, and Grabowski R, 2018. Factors influencing knowledge of HIV/AIDS in Nepal: role of socioeconomic interactions. Journal of Social and Economic Development, 20(1), pp.17-191.

[15] Banerjee A, La Ferrara E, and Orozco-Olvera VH, 2019. The entertaining way to behavioral change: Fighting HIV with MTV. The World Bank.

[16] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[17] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[18] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[19] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[20] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[21] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[22] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[23] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[24] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[25] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[26] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[27] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[28] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[29] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[30] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[31] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[32] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[33] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[34] Siril HN, Kaaya SF, Fawzi MKS, Mtisi E, Somba M, Kilewo J, et al. CLINICAL outcomes and loss to follow-up among people living with HIV participating in the NAMWEZA intervention in Dar es Salaam, Tanzania: a prospective cohort study. AIDS research and therapy. 2017;14(1):18.

[35] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[36] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[37] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[38] Peltzer K, Matseke G, Mzolo T, Majaja M. Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. BMC public health. 2009;9(1):174.

[39] Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry. 2009;8(1):50-61.

[40] Easterly W. How the millennium development goals are unfair to Africa. World development. 2009;37(1):26–35.

[41] Barbarin OA, Khomo N. Indicators of economic status and social capital in South African townships: What do they reveal about the material and social conditions in families of poor children? Childhood. 1997;4(2):193–222.

[42] Smith JA, McPherson M, Smith-Lovin L. Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. American Sociological Review. 2014 Jun;79(3):432-56.

[43] Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to censor? American journal of epidemiology. 2017;187(3):623–632.

[44] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.). 2010;21(1):13.

[45] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[46] Hong, G., 2015. Causality in a social world: Moderation, mediation and spill-over. John Wiley & Sons.

[47] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[48] Imai K, Jiang Z. Identification and sensitivity analysis of contagion effects in randomized placebo-controlled trials. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2019.

[49] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR.2012;40(2):211–39.

[50] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

[51] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[52] Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013;18(2):137.

[53] Leung MP. Treatment and spillover effects under network interference. Review of Economics and Statistics. 2016;p. 1–42.

[54] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[55] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

[56] Berkman LF. Social support, social networks, social cohesion and health. Social work in health care. 2000 Sep 26;31(2):3-14.

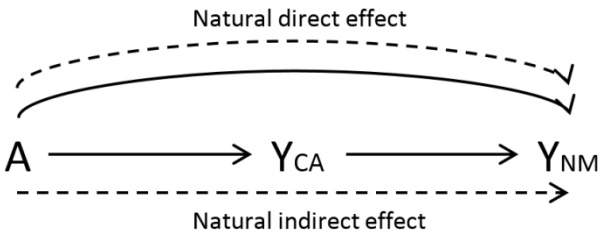
[57] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

[58] Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Social science & medicine. 2015;125:192–202.

[59] He P, Eriksson F, Scheike TH, Zhang MJ. A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. Scandinavian Journal of Statistics. 2016 Mar;43(1):103-22.

11. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE). The NDE indicates the increase in NMs' HIV knowledge happens without a concomitant increase in their CA's HIV knowledge. The NIE, on the other hand, indicates that the increase in NMs' HIV knowledge is mediated by their CA's HIV knowledge increasing. Solid lines indicate paths of causality between variables. Dashed lines represent the line or lines composing the effect of interest.



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

BMJ Open

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033759.R5
Article Type:	Original research
Date Submitted by the Author:	06-Aug-2020
Complete List of Authors:	Lienert, Jeffrey; University of Pennsylvania Perelman School of Medicine, Center for Health Care Innovation; Philadelphia VAMC, CHERP Fawzi, Mary C; Harvard Medical School, Global Health and Social Medicine McAdam, Keith; London School of Hygeine and Tropical Medicine, Epidemiology Kaaya, Sylvia; Muhimbili University of Health and Allied Sciences, School of Medicine Liu, Yuanyuan Todd, Jim; London School of Hygeine and Tropical Medicine, Population Health Andrew, Irene; Management and Development for Health, Onnela, Jukka Pekka; Harvard T.H. Chan School of Public Health, Biostatistics
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Global health, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating spillover of HIV knowledge from study participants to their network members in a stepped-wedge behavioral intervention in Tanzania

Jeffrey Lienert[a,b,j], Mary C. Smith Fawzi[c], Keith McAdam[d], Sylvia Kaaya[e], Yuanyuan Liu[i], Jim Todd[g], Irene Andrew Lema[f], Jukka-Pekka Onnela[h]

[a]*Perelman School of Medicine, University of Pennsylvania* [b]*Center for Health Equity Research and Promotion, Philadelphia VA* [c]*Department of Global Health and Social Medicine, Harvard Medical School* [d]*Department of Clinical Tropical Medicine, London School of Hygiene and Tropical Medicine* [e]*Department of Psychiatry and Mental Health, Muhimbili University of Health and Allied Sciences* [f]*Management and Development for Health* [g]*Department of Population Health, London School of Hygiene and Tropical Medicine* [h]*Department of Biostatistics, Harvard T. H. Chan School of Public Health* [i]*University of Texas School of Public Health*
[j]*Corresponding Author: Jeffrey Lienert, jeffrey.lienert@pennmedicine.upenn.edu*

Preprint submitted to BMJ Open *3 August, 2020*

Abstract

Objectives: We aim to describe the social network members of participants of a behavioral intervention, and examine how the effects of the intervention may spillover among network members.

Design: Secondary analysis of a step-wedge randomized controlled trial.

Setting: Change Agents (CAs) were recruited from waiting rooms of HIV treatment facilities in Dar es Salaam, Tanzania, and their Network Members (NMs) were recruited directly by CAs.

Participants: We enrolled 662 CAs in an HIV behavioral intervention. They, along with 710 of their NMs, completed baseline and follow-up interviews from 2011 to 2013.

Primary and Secondary Outcomes: The primary outcome of this study was change in NMs' HIV knowledge, and the secondary outcome was whether the NM was lost to follow-up.

Results: At baseline, many characteristics were different between NMs and CAs. We found a number of NM characteristics significantly associated with follow-up of NMs, particularly female gender (OR=1.64, 95% CI: 1.02,2.63) and HIV knowledge (OR=20.0, 95% CI: 3.70,125); only one CA variable was significantly associated with NM follow-up: having a private source of water (OR=2.17, 95% CI: 1.33,3.57). The 14.2% increase in NMs' HIV knowledge was largely due to CAs feeling empowered to pass on prior knowledge, rather than transmitting new knowledge to their NMs.

Conclusions: Characteristics of social network members of persons living with HIV PLH may play a role in study retention. Additionally, the HIV knowledge of these NMs increased largely as a function of CA participation in the intervention, suggesting that intervening among highly-connected individuals may maximize benefits to the potential population for whom spillover can occur.

Keywords:

mediation, natural direct effect, natural indirect effect, follow-up, retention, hidden population

1. Strengths and Limitations of This Study

- The recruitment method, focused on social network members of persons living with HIV, allowed us to access an at-risk population typically not easily-accessible via other means, such as direct approach by the study team rather than someone known to the participant.
- The study design of the trial in which this study is nested was an ideal situation in which to employ a mediation analysis that informs our understanding of how this and similar intervention effects may spillover in a population.
- The greater-than-ideal dropout rate of NMs was both a strength in that it allowed us to examine factors associated with dropout, but also a limitation, in that the potential of differential dropout by unmeasured factors may have biased some of our results.

2. Introduction

Despite recent decreases in mortality, HIV/AIDS is currently the ninth leading cause of years of life lost (YLL) globally [1]. Research into eliminating HIV has focused on two fronts: biological and behavioral, which combined have decreased the years of life lost due to HIV by an estimated 51% [1]. Behavioral interventions seek to curb transmission of HIV through reduction of risk behavior including safer sexual practices and reducing injection drug use [2, 3]. Behavioral change also impacts the effectiveness of the biological methods, given that lower adherence is known to often compromise their effectiveness [4]. However, neither biological nor behavioral methods are fully effective without the knowledge of these treatments and preventative behaviors. It is therefore important to increase HIV-related knowledge as a necessary, but not sufficient, step towards achieving the 90-90-90 treatment goals (90% diagnosis, 90% antiretroviral therapy, and 90% viral suppression among the treated) and ultimately eliminate HIV [5].

The importance of adequate HIV knowledge has been recognized for a period of time, as it is often a prerequisite for behavioral change [6]. It should be noted that increased knowledge does not necessarily translate into behavior change, as other factors such as motivation and skills also play a role [6]. Despite this, researchers conducted trials among Persons Living with HIV (PLH) which led to increased HIV knowledge [7, 8]. These interventions were group discussions and teaching led by nurses, and motivational

interviewing, respectively. Researchers also found that the intervention reduced HIV-risk behaviors concomitantly with an increase in HIV knowledge [8].

1. Knowledge gained by participants in these trials can also be freely shared with members of their social network, in what is known as a spillover effect [9]. Specifically, a spillover effect (also known as an indirect or disseminated effect) is one person's exposure affecting another's outcome [10]. In the context of a social network spillover intervention, it corresponds to an individual who was unexposed to an intervention changing their behavior because they were socially-connected to an individual who did receive the intervention. This is distinct from what is sometimes called behavioral spillover, where changes in a person's behavior affect other behaviors of that same person [11]. For injection drug users, HIV prevention educational interventions were demonstrated to have spillover effects of HIV prevention education, and subsequent reduced rates of risky behaviors [12]. Studies have also used proxy variables for social network ties such as inviting social network members to watch educational programming [13] or time spent shopping at the market [14] to evaluate spillover effects for HIV knowledge, generally finding evidence for spillover. However, spillover in HIV knowledge between known social network ties generally remains understudied, particularly in sub-Saharan Africa [15]. We therefore aim to determine whether social network members those receiving an HIV behavioral/knowledge intervention also increase their HIV knowledge.

New knowledge can come from a variety of sources, one of the most important of which is a person's social network [16, 17]. Social networks are of particular import because new knowledge can lead to cascades of behavior change, where people subsequently educate those in their social network, in what is known as social influence [18, 19, 20]. This has been examined in participant-driven interventions, where initially-recruited participants educate members of their social network one-on-one [21]. Characteristics such as knowing the HIV status of network members has been shown to be the most important predictor of engaging in prevention advocacy [22]. Work on diffusion through social networks, how a belief or behavior can be "contagious" within a network, has shown that spreading intervention effects beyond the initial study population can improve the cost-effectiveness of these interventions [23]. These findings imply that certain aspects of knowledge or behavior may spread more or less efficiently through networks comprising individuals with specific characteristics, which may need to be accounted for in network interventions. For instance, networks comprising many at-risk individuals who are HIV-negative may not be as receptive to change in behaviors as networks comprising a mix of those with and without HIV. Additionally, the networks of PLH are often difficult networks to ascertain [24], due to the continued stigma of HIV and AIDS

in many settings [25]. Because of this, if a PLH or person at-risk of acquiring HIV does not want to participate in an investigator-initiated intervention, there is little recourse other than information transmitted via social networks, or targeted sampling techniques which are not always effective (e.g. Respondent Driven Sampling) [26]. This is particularly important in low- and middle-income countries, as a recent systematic review found only 54 studies researching spillover effects in these settings (out of approximately 750) [27]. Therefore, understanding exactly how information spreads from participants in an intervention to members of their social network, who may be largely inaccessible via other means, is important for reaching the greatest number of people about HIV prevention. Understanding what makes these persons different from those who receive the intervention themselves is important, as it may point to ways in which to increase enrollment of these populations.

Based on the above gaps in the literature, we conduct a study on network members of PLH enrolled in a behavior change intervention [28]. The trial recruited PLH to serve as Change Agents (CAs) and to reach out to their social network members (NMs) about knowledge of HIV and safer sexual practices [29]. Our goals in this study were threefold: 1) to understand how the NMs differed from their CAs, 2) understand correlates of dropout for the NMs, and 3) to understand by what method HIV knowledge transferred to the NMs from the CAs. Understanding how the information and behaviors are shared within social networks will allow HIV researchers and others to take advantage of this knowledge and improve upon prevention interventions in the future.

3. Methods

13.1.Study Population

We analyze social network data from the Agents of Change trial [30], which was a stepped-wedge randomized controlled trial [31] that enrolled PLH to become Change Agents (CAs) by informing members of their social network (NMs) about knowledge of HIV and safer sexual practices. Here, we define CAs based on the *potential* for PLH to become so - by self-selecting into the study, PLH identify themselves as potential CAs. Although we refer to them as 'CAs' throughout, participants in the trial enrolled with varying levels of ability to act as a Change Agent. Through receiving the *NAMWEZA* intervention, we hypothesize that CAs will be able to truly self-actualize and subsequently act as Change Agents in their community.

CAs were recruited from the waiting rooms of HIV care and treatment centers in Dar es Salaam, Tanzania, and we received written consent from each CA. Participants completed a baseline questionnaire and

were randomized to one of three waves in which to receive the intervention. At baseline, participants were also asked to recruit up to three members of their social networks who they felt were at particularly high risk of contracting or spreading HIV. We obtained written consent from these nominated network members (NMs). NMs could be either HIV positive or negative, and they were given a baseline survey. The NM was only aware the CA was a participant in the intervention if the CA shared this information with them, which many did not due to HIV-related stigma [32]. Each CA therefore formed a CA-NM dyad with each NM they recruited, and if they recruited more than one NM, formed a set of CA-NM dyads with a common CA.

As fits a stepped-wedge RCT, all CAs eventually received the intervention, but were randomized to *when* they received it. These waves each lasted 12 weeks, at which point the next wave began and another group of CAs received the intervention. Within each wave, the intervention comprised 10 weekly structured sessions aimed at empowering PLH to become HIV prevention change agents in their communities. The sessions aimed to increase psychosocial and communication skills through an Appreciative Inquiry approach [33]. Within one month of each wave of the intervention, CAs were given follow-up surveys. Across all waves, the interventions lasted from November 2010 to January 2014, and the final interviews were conducted in March 2014. For more information on the study design, we direct interested readers to Smith-Fawzi et al., 2019 [29].

The NMs did not receive any intervention at any point during the study. Rather, their intervention status flowed from the intervention status of their CAs. Due to operational difficulties, each NM was surveyed two times during the study: baseline and after the first wave, rather than baseline and one after each wave. In this way, all demographic and contextual variables were measured at baseline. We use this interview of NMs as the division between "exposed" and "unexposed". At the time of an NM's follow-up interview, their respective CA may or may not have undergone the intervention. In other words, the CAs randomized to receive *NAMWEZA* during the first wave would have potentially indirectly exposed their NMs to the intervention when the NM completed their follow-up questionnaire after Wave 1. Therefore, the NMs were divided into "exposed" (N=381) and "unexposed" (N=329) groups based on whether their respective CA was randomized into receiving *NAMWEZA* during the first wave or not. The CAs always completed their Wave 1 follow-up interview before their NMs were invited to complete their Wave 1 follow-up interview.

Inclusion criteria for the follow-up analyses were for CAs who completed at least one follow-up interview and had at least one NM who completed at least one follow-up questionnaire. In this way, outcomes could be computed for the CA and NM of each dyad. NMs approached by a CA, but who did not participate

in the study were not recorded since it was not feasible to obtain this information from study participants themselves. As we lost some CAs and NMs to follow up, we completed our analyses without their data, assuming it to be Missing Completely at Random (MCAR). During this study, there was little loss-to-follow-up among the CAs (< 10%), but much higher loss among the NMs (36.8%) [34]. Given an NM or CA was not lost to follow-up, complete information was available on all additional variables, including exposure, outcome, and covariates. In sum, our sample comprises 662 CAs and 710 NMs, meaning each CA recruited 1.07 NMs on average out of a possible 3, and 44 CAs nominated at least two NMs.

13.2.HIV knowledge

To assess HIV knowledge of CAs and NMs, we used the brief HIV knowledge questionnaire [35]. This scale comprises 18 questions, which focus on an individual's knowledge of how HIV can spread and other characteristics of the virus and AIDS. This knowledge is crucial for an individual's subsequent safe-sex practices to reduce the risk of transmission. The original population comprised three different groups: two groups of low-income women, and one of women and men receiving psychiatric treatment. In these populations, questions on the measure had a Chronbach's alpha of 0.78. This instrument has been used previously in sub-Saharan Africa and has demonstrated reasonable reliability of 0.75 Cronbach's alpha in South Africa among a convenience sample of 429 members of the African Methodist Episcopal church [36]. It has also been translated to Swahili, with only minor differential item functioning [37]. This indicates that the measure performs adequately in other, similar populations. In the present study the Cronbach's alpha is 0.71. Because the average baseline score of HIV knowledge was 0.80, with a preponderance of scores of 1.00, the main indicator of knowledge was not normally distributed, and therefore a continuous predictor was not ideal [38,39]. We therefore summarize this measure as "Complete HIV Knowledge", a dichotomous variable for whether the participant correctly answered all questions.

13.3.Demographic and contextual variables

In addition to HIV knowledge, we are interested in a set of demographic and contextual variables that may help explain some of the trends observed. In terms of demographic variables, we include age, sex, employment status, marital status, and self-identified HIV status. We also include education, which we dichotomize at 7 years or more - a level commensurate with elementary education. We selected this cutoff

because it coincides with the millennium development goal (MDG) of increasing primary education completion [40].

Contextually, we included two additional variables: having a private source of water, and the number of persons sleeping in the participant's home. The first was also based on a MDG, and indicates participants with access to safe drinking water [40]. This is a proxy for the economic security of the participant. The number of persons sleeping in the participant's home is also a non-monetary indicator of their material and social resources [41]. These variables combined give a more thorough picture of the participant's economic status than employment alone.

13.4.Statistical Analysis

The first analysis was to assess homophily, defined as the extent to which CAs and NMs were similar to one another in terms of a number of sociodemographic characteristics and HIV-related risk factors. Because CAs and NMs self-selected into their respective group, and only CAs were randomized (with their NMs being randomized along with them), differences between the groups were to be expected. However, we only examined homophily of baseline characteristics rather than of outcome, because comparison of outcomes between CAs and NMs would remove the benefits of randomization. Because CAs and NMs self-selected into the study and were not randomized to CA/NM status, we do not *a priori* expect them to be completely similar. In addition, multiple NMs could share a CA and would therefore not be independent due to the shared variation and latent characteristics of having the same CA. We therefore assessed statistical significance of homophily on the set of CA-NM dyads using a permutation test, a non-parametric test which has no distributional assumptions. For continuous variables, the difference between the CA and NM was calculated, and for categorical variables, whether the CA and NM were concordant or discordant was recorded. We then randomly permuted CA-NM ties (keeping number of ties per CA constant), and then recalculated the difference or percent concordant, respectively, 1,000 times. We then examine the percentile of the observed difference relative to the permuted differences [42]. Analyses were run using R v3.1.1.

To accomplish our second aim of understanding what was associated with NMs completing their follow-up interview, we fit a log-binomial regression to determine predictors of follow-up. In this regression, we use all the variables listed above, as well as whether the CA remained in the study for its full duration. In order to examine the association between these same variables and the time of follow-up, we also employed a Cox regression. In the Cox regression, start of follow-up was defined as the time at which the NM

completed their baseline interview. The outcome here was whether the NM completed a follow-up interview. NMs who were lost to follow-up were censored at the time of their latest interview [43,44].

Finally, since the trial showed beneficial effects on the HIV knowledge of the NMs following the intervention [30], we aim to elucidate exactly what caused the HIV knowledge of NMs to increase - either CAs gaining knowledge through the intervention and sharing it, or the CAs being empowered by the intervention to share existing knowledge. As the wedge in which the CA received the *NAMWEZA* intervention was randomized, we treat each NM as being randomized to exposure to *NAMWEZA* at the same time as their CA. This randomization scheme allows us to explore the spillover effect of CAs' HIV knowledge onto their respective NMs via a mediation analysis. These pathways represent different types of spillover effects: the exposure or outcome of one person affecting the outcome of another person.

As shown by VanderWeele et al (2015), social network spillover effects in the case of dyadic relationships can be broken down into concepts from mediation analysis: direct and indirect effects (Figure 1) [45]. This method has since been used for novel evaluations of spillover effects [46,47,48]. Although previous studies showed that this same type of analysis cannot be done on full network data, the data in this study consisted of only dyads, the CAs and their NMs, so in this case the method is appropriate [49,50].

The method parses social influence into direct and indirect effects. The natural indirect effect (NIE) is the effect by which the intervention changes the CA's HIV knowledge, which then changes the NM's HIV knowledge. The NIE is similar to the effect observed in many participant-driven interventions: an initial participant receives the intervention, increasing their knowledge, and they subsequently pass their increased knowledge to members of their social network [21]. The natural direct effect (NDE) is the effect of receiving the intervention has on an NMs outcome, irrespective of the CA's outcome (in this case HIV knowledge). For instance, this could occur if CAs begin with good knowledge of HIV, and the intervention empowers them to convey knowledge they already had to their NMs. Although the intervention does not increase their HIV knowledge, it is still useful to the CAs, as it empowers them to act as CAs in their community. In order to estimate these effects, the published SAS macro developed by VanderWeele and colleagues was used, and analyses were run using SAS v9.2 (SAS Institute, Cary, NC) [51,52]. In the models estimating the effect of the exposure on the mediator and estimating the effect of the mediator on the outcome, we adjusted for all the variables included in our logistic regression above. Although the randomization of the exposure minimized some potential bias, the loss-to-follow-up among the

CAs indicates that selection bias could remain a concern, so we control for the variables which may also impact loss-to-follow-up.

Importantly, this analysis requires a number of assumptions and applies to dyads only when these assumptions are met. One assumption of this analysis is that the dyads are independent, which is violated here; if a CA recruited more than one NM, the multiple CA-NM dyads involving the same CA would not be independent. To address this, we performed the analysis after randomly removing NMs until each CA had only a single NM. This resulted in removing 48 NMs, just 6.7% of the population. We found that the point estimates were nearly identical, but that the confidence intervals were slightly larger due to the reduced sample size. No coefficients changed from significant to non-significant in this analysis (data not shown).

A second, related assumption is that of *partial interference*, that the effects in one cluster does not affect another cluster - here, one CA-NM dyad affecting another [53]. This could occur if two NMs of different CAs happen to know one-another outside of the study, one has a CA who was randomized to an earlier wedge, and shares what they know of it with the otherwise-unexposed NM. However, due to the large size of Dar es Salaam, and the number of HIV treatment clinics in which recruitment occurred, we expect few CAs or NMs to know one another outside of the study (other than NMs knowing the CA who recruited them), limiting the potential for partial interference.

A third assumption of this analysis is that the outcomes of the CA and NM are independent conditional on the CA's exposure, or conditional on the CA's exposure and any confounding variables [45,46]. Because CA-NM pairs self-select and are not randomized, we do not expect these outcomes to be independent conditional of the CA's exposure, and so we adjust for additional variables to meet this assumption. In our analysis, we therefore adjust for all the variables used in the log-binomial regression.

The data are not publicly available due to the sensitive nature of HIV infection status and the socially-networked nature of the data. Because the data include specific information on social ties, some of whom have not disclosed HIV sero-status to one another, the risks of individual identification and compromising HIV sero-status are greatly increased.

13.5.Patient and Public Involvement

Patients were not involved in the design of this study. As part of the recruitment process, patients were instructed to recruit members of their social networks who they felt were at high risk of HIV infection - in

this way participants could guide the recruitment of NMs to those most at-risk. We currently have no plans to disseminate the results of this study to participants.

4. Results

The *NAMWEZA* study recruited 662 Change agents (CA), of whom 453 (68.4%) completed at least one follow-up questionnaire. These 662 CAs in turn recruited 710 network members (NM) who took the baseline questionnaire. Of the 710 NMs, 449 (63.2%) also completed a follow-up questionnaire (Table 1). At baseline, CAs were on average older than their NMs. More of the NMs were employed than their CAs (69.3% vs. 55.0%, p<0.001), but were less likely to have at least 7 years of education (52.0% vs. 52.3%, p<0.001). Only 12.3% of NMs were HIV-positive, compared to *all* CAs (p<0.001). Complete data was obtained at baseline for all CAs and NMs.

Characteristic	Number of NMs (%) or Mean (SD) (N=710)	Number of CAs (%) or Mean (SD) (N=662)	P-value
Age	33.0 (11.1)	38.9 (9.7)	<0.001
Female	380 (53.7%)	349 (53.9%)	0.89
Employed	490 (69.3%)	356 (55.0%)	<0.001
At least 7 years education	369 (52.0%)	584 (82.3%)	<0.001
Complete HIV knowledge	638 (89.9%)	598 (90.4%)	0.65
Persons sleeping in home	5.00 (5.77)	3.87 (3.72)	<0.001
Married	373 (52.7%)	338 (51.1%)	0.56
HIV Positive	87 (12.3%)	662 (100%)	N/A
Private source of water	309 (43.7%)	263 (39.7%)	0.19

Table 1: Demographic characteristics at baseline, with the results of a permutation test for homophily. Specifically, for each CA- NM dyad, either a difference (for continuous variables) or concordance (for dichotomous variables) is calculated. For example, if a CA was 39 years old, and their NM was 25 years old, the difference would be 14 years old. If a CA was male and their NM was Female, the pair would be discordant for sex. CA-NM pairs were then randomly reshuffled, the edge-wise characteristics recalculated, and the observed difference compared to the distribution of randomized differences.

Risk ratios (RRs) obtained via log-binomial regression indicated that characteristics of both the CAs and the NMs significantly predicted loss to follow-up (Table 2). The NM being female (RR=1.44, 95% CI: 1.05,1.97), having complete HIV knowledge (RR=10, 95% CI: 2.33,42), being employed (RR=1.43, 95% CI: 1.08,1.89), and being married (RR=1.55, 95% CI: 1.03,2.33) were all significant predictors of increased odds of completing a follow-up interview. Each additional person sleeping in the home of the NM per room used for sleeping reduced the odds of follow-up (RR=0.85, 95% CI: 0.74,0.98) as did the NM living with

HIV (RR=0.40, 95% CI: 0.17,0.96). CA having a private water source was significantly associated with increased odds of the NM being followed-up (RR=2.07, 95% CI: 1.25,3.42), even after controlling for the NM having a private source of water. This was the only CA-specific variable that significantly predicted an NM's follow-up.

The Cox proportional hazard model showed similar results to the logistic regression, but with fewer significant results. Only the NM being married (HR=1.28, 95% CI: 1.01,1.64) and the NM living with HIV (HR=0.71, 95% CI: 0.51,0.99) significantly predicted the time-to-follow-up of the NMs. Additionally, all of the hazard ratios were closer to the null relative to the corresponding odds ratios.

We assess the mechanism of spillover effects on HIV knowledge, the main effect of the intervention on the NMs. The analysis showed that NMs who were exposed to intervention via their CA had an increase in HIV knowledge of 12.9% on average (95% CI: 0.06, 0.20) from a baseline percentage of 78% [26]. This was broken down into a natural indirect effect (NIE) of 0.6% (95% CI: -0.06%,2.0%), which is the effect the intervention had via the CA's HIV knowledge, and a natural direct effect (NDE) of 12.3% (95% CI: 6.1%,19.3%), which is the effect the CA's participation had on the NM's HIV knowledge, irrespective of the CA's HIV knowledge. In other words, of the 12.9% increase in NM's having complete HIV knowledge, 12.3% occurred without a concomitant increase in their CA's HIV knowledge. In other words, their HIV knowledge increase was not mediated by the increase in HIV knowledge of their CA. This did not change when we used only one NM per CA.

5. Discussion

Based on results from the NAMWEZA trial in Tanzania, we have shown a number of novel findings regarding the network members of HIV intervention participants, correlates of their retention in the study, and the evidence for spillover of HIV knowledge from CAs to NMs. These findings can inform the design of interventions in the future to maximally enroll and retain participants and ensure that intervention information is transmitted from the study participants to members of their social networks.

Characteristic (N=459)	Adjusted RR (95% CI)	Adjusted HR (95% CI)
Characteristics of NMs		
Gender (Female)	1.44* (1.05,1.97)	1.18 (0.94,1.50)
Difference in age of NM and CA (per year)	1.01 (0.99,1.03)	1.01 (0.995,1.02)
Age of NM	0.98 (0.95,1.02)	0.99 (0.97,1.04)
Complete HIV knowledge (vs. none)	10* (2.33,42)	2.20 (0.97,5.01)
Employed	1.43* (1.08,1.89)	1.15 (0.89,1.50)
Each additional person sleeping in home per room used for sleeping	0.85* (0.74,0.98)	0.92 (0.83,1.01)

Married	1.55* (1.03,2.33)	1.28* (1.01,1.64)
Living with HIV	0.40* (0.17,0.96)	0.71* (0.51,0.99)
Having a private source of water	0.89 (0.59,1.34)	0.97 (0.77,1.22)
Characteristics of CAs		
Gender (Female)	1.27 (0.76,2.08)	1.07 (0.84,1.37)
Complete HIV knowledge (vs. none)	0.36 (0.07,2.04)	0.64 (0.29,1.43)
Having a private source of water	2.07* (1.25,3.42)	0.97 (0.77,1.22)
Being employed	1.54 (0.95,2.50)	1.14 (0.90,1.43)
Being married	1.25 (0.77,2.04)	1.11 (0.88,1.43)
CA Lost to follow-up	1.06 (0.84,1.33)	1.02 (0.72,1.43)

Table 2: Results of multivariate log-binomial regression and Cox-proportional hazard models on the dichotomous outcomes of whether the NM completed a follow-up questionnaire, and the continuous outcome of time-to-completion of follow-up questionnaire, respectively. "*" indicates significant at the $p < 0.05$ level.

We found several significant differences between CAs and NMs they recruited. On average, CAs were older, less likely to be employed, and more educated. All of this suggests that by having CAs recruit from their social network, we were able to recruit a set of social network members different in multiple ways from those found in the waiting rooms in HIV treatment clinics (unsurprising, given that many NMs were not HIV-positive) [9]. This may in part be due to the fact that when recruiting NMs, CAs were instructed to recruit those who may be at particular risk of contracting or spreading HIV. Therefore, CAs likely did not pick random members of their social network, but those who were at high risk. This finding corroborates the claim by Latkin et al. (2013) that sampling via enumeration of network members by initial participants recruits a more diverse sample, as our sample of NMs was not composed of only those who were HIV-positive, but also many who were HIV-negative [54]. Additionally, the lower likelihood of employment of the CAs indicates they may have had more time for participating in the intervention [55]. This means that future studies may need to tailor their interventions to work with the schedules of employed persons to increase participation.

Although the study design potentially accessed a separate slice of the population than other methods vis-a´-vis the population of those at risk for contracting HIV, it may have been at least partially responsible for the large loss of follow-up of the NMs. It is important to note that more NM characteristics than CA characteristics were significantly predictive of follow-up in both the logistic and Cox proportional hazards models. The one CA characteristic which did predict NMs completing a follow-up interview, having a private source of water, was a proxy for CA's socioeconomic status. This may have been because CAs with greater resources may have had more time available to pass information to their NMs, retaining the NM's interest longer [56]. We also note that less HIV knowledge and living with HIV predicted reduced likelihood

of completing a follow-up questionnaire among NMs, which means that those who might have benefited most from spillover of the intervention were more likely to discontinue their involvement. This does not mean that they did not receive any spillover, only that it was not recorded. Our estimate of the magnitude of the spillover may therefore be biased towards the null. This is also problematic more generally for interventions of this nature as the very people the intervention aims to benefit may not stay with the program.

The HIV knowledge gain experienced by the NMs was largely due to the NDE; i.e. knowledge spilled-over as a result of the CAs participating in NAMWEZA, not because the CAs' HIV knowledge increased, and they passed this new knowledge to their NM(s). This is most likely because CAs had a high average HIV knowledge score at baseline (80%), so there was less room for improvements in knowledge. Following the intervention, CAs' likelihood of complete HIV knowledge did not significantly increase. Therefore, what prompted the increased knowledge of the NMs was the CA becoming empowered through the intervention to pass on their existing knowledge to their NM [57].

This finding is important for future interventions: spillover effects of this intervention will likely carry over only to those directly-connected to the CAs (opposed to spreading indefinitely in a snowball effect), as the most important thing in transmitting HIV knowledge is receiving the intervention, which only CAs receive. Because NMs do not actually receive the intervention, it is unlikely that they would feel empowered to become CAs themselves (particularly because they will not receive the *NAMWEZA* sessions), thus limiting the spread of the intervention. For CAs to increase the HIV knowledge of their NMs, they would only need to become empowered to share their information. However, for the NMs to increase the HIV knowledge of their own NMs, they would need to both increase their HIV knowledge, and become empowered to share it. This is less likely to happen than just becoming empowered, and so it is unlikely, though not impossible, that this effect would continue to spread in the population. This may give insight into how to design interventions in the future; if one wants to maximize the number of people who benefit, choosing CAs who form many bridging ties in the community would maximize the potential number of links by which spillover can occur [58]. Alternatively, interventions can be designed to be self-propagating; if CAs are empowered to deliver the intervention to others, changing their own NMs into future CAs, those new NMs-turned-CAs could then deliver the intervention to a second set of NMs, again empowering them to become CAs.

Our study is not without limitations. First, there was considerable loss to follow-up of the NMs. Although this actually informed our analysis of the correlates of loss to follow-up, it meant that our analysis of spillover effects may be biased. Even though the exposure was randomized, the loss to follow-up can

result in selection bias if the NMs who left the study were systematically different from those who remained. As we show, the NMs who dropped out were those who would have benefited the most from the intervention because they were more vulnerable and at higher risk than those who did not drop out. Because they are likely to have greater benefits from any spillover effects, we would expect this non-random loss to follow-up to result in an underestimate of the impact of the intervention in the network members. Adjusting for censoring weights may ameliorate this issue [59]. Second, our use of Hazard Ratios (HRs) has important limitations: they are subject to selection bias, are sensitive to study period, and only provide one estimate during the study [44]. Any of these limitations could affect this analysis, hence our use of logistic regression as a primary analysis. However, they remain useful as a sensitivity analysis. Third, our data did not perfectly fit the requirements of the causal mediation analysis: they were not entirely independent, since multiple NM-CA dyads shared a CA. However, when we randomly removed dyads until there were no repeated CAs, the results were qualitatively very similar, indicating that lack of independence did not unduly affect our results. Fourth, although we were able to tease apart the direct and indirect effects, we are unable to determine the mechanism of the natural direct effect; the data do not allow us to specify whether NMs increased knowledge through speaking to their knowledgeable CA, through researching HIV on their own, or some other mechanism. Future work will have to be done to examine these different pathways.

6. Conclusions

These results have implications for the potential scale-up of the *NAMWEZA* intervention, as well as future studies and interventions that focus on behavioral interventions in social networks. First, our findings of minimal similarity between CAs and their NMs indicate that this recruitment method allows us to enroll participants from portions of the population that are not represented by the CAs alone. Coupling this with a deeper understanding of the mental heuristics CAs used to select NMs (e.g. did CAs mentally search their close or peripheral network for those at-risk of HIV), may lead to different strategies for recruitment and retention, leading to stronger effects of behavioral interventions. The mediation analysis presents a compelling picture of how best to ensure the benefits of interventions reach as many people beyond the study participants as possible. Participation of CAs in the intervention resulted in positive effects on their immediate network members' HIV knowledge regardless of how the CAs responded to the intervention. While improvement in HIV knowledge may not necessarily translate to increased safe sex behavior, it can be seen as a rate-limiting step towards reducing HIV transmission since adequate knowledge is usually

necessary for reducing risk behavior. Future work should examine the exact mechanisms of spillover discussed here, as that is an important clarification that could benefit future studies. Specifically, a similar setup to the study here combined with semi-structured interviews with CAs and NMs about their interactions with one another would help elucidate exactly how NMs increased their HIV knowledge (this population would likely need to comprise CAs who have disclosed their HIV status in order to prevent accidental disclosure). Future work should also examine whether increases in HIV knowledge translate to changes in behaviors which may increase one's risk of contracting HIV, particularly in Sub-Saharan Africa. The results presented herein may inform approaches for increasing participation and potentially conferring greater benefits related to spillover effects in future HIV behavioral interventions.

7. Acknowledgements

The authors would like to thank the study participants and staff for their time, hard work, and dedication. We would also like to thank Guy Harling for their comments on the manuscript.

8. Contributorship

JL, MCSF, and JPO designed the analysis plan and drafted the initial paper. JPL, MCSF, KM, SK, IAL and JT conducted the underlying randomized trial, and designed it to incorporate social networks data and spillover information; these same authors provided feedback on the proposed analysis plan. JL and YL conducted analyses for the paper. All authors helped revise the drafts of the paper and approved the final manuscript.

9. Compliance with ethical standards

Funding: The authors received funding from the CDC via grant TZ/UG.08.0147, the President's Emergency Plan for AIDS Relief - Tanzania, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant RP160015).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. IRB approval was obtained from Harvard Medical School, the Muhimbili University of Health and Allied Sciences, and the National Institute

for Medical Research (NIMR) in Tanzania. Particular attention was paid to eliminate the risk of inadvertent disclosure of CA HIV status to NMs.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Data sharing statement: There are no additional data available. As the data contain enough information to potentially-uniquely identify specific participants, in the sensitive context of HIV, we have chosen not to make the data available.

10. References

[1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–1210.

[2] Crepaz N, Tungol-Ashmon MV, Higa DH, Vosburgh W, Mullins MM, Barham T, et al. A systematic review of interventions for reducing HIV risk behaviors among people living with HIV in the United States, 1988-2012. AIDS. 2014 mar;28(5):633–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24983541.

[3] Arnold EA, Kegeles SM, Pollack LM, Neilands TB, Cornwell SM, Stewart WR, et al. A Randomized Controlled Trial to Reduce HIV-Related Risk in African American Men Who Have Sex with Men and Women: the Bruthas Project. Prevention Science. 2018;p. 1–11.

[4] Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes J, Cayla J. Impact of adherence and highly active retroviral therapy on survival in HIV-infected patients. Journal of Acquired Immune Deificiency Syndromes. 2002;30(1):105–10.

[5] Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids and others. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS. 2014;.

[6] Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological bulletin. 1992;111(3):455.

[7] McNeill C, George N, Glover R. An Evaluation of Sisters Informing Healing Living Empowering: Increasing HIV Knowledge Among African American Adolescent Females Using an Evidence-Based HIV Prevention Intervention. Journal of Doctoral Nursing Practice. 2017;10(1):4–10.

[8] Carey MP, Braaten LS, Maisto SA, Gleason JR, Forsyth AD, Durant LE, et al. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology. 2000;19(1):3.

[9] Sinclair B, McConnell M, Green DP. Detecting spillover effects: Design and analysis of multilevel experiments. American Journal of Political Science. 2012;56(4):1055–1069.

[10] Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford Jr JM, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International journal of epidemiology. 2017;47(1):332–347.

[11] Truelove HB, Carrico AR, Weber EU, Raimi KT, Vandenbergh MP. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environmental Change. 2014;29:127–138.

[12] Ghosh D, Krishnan A, Gibson B, Brown SE, Latkin CA, Altice FL. Social network strategies to address HIV prevention and treatment continuum of care among at-risk and HIV-infected substance users: a systematic scoping review. AIDS and Behavior. 2017;21(4):1183–1207.

[13] Benjamin-Chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, Tran D, Arnold BF, Hubbard AE, Luby SP, and Miguel E, 2017. Spillover effects on health outcomes in low-and middle-income countries: a systematic review. International journal of epidemiology, 46(4), pp.1251-1276.

[14] Self S, and Grabowski R, 2018. Factors influencing knowledge of HIV/AIDS in Nepal: role of socioeconomic interactions. Journal of Social and Economic Development, 20(1), pp.17-191.

[15] Banerjee A, La Ferrara E, and Orozco-Olvera VH, 2019. The entertaining way to behavioral change: Fighting HIV with MTV. The World Bank.

[16] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. The New England journal of medicine. 2007;357(4):370–379.

[17] Perkins JM, Subramanian SV, Christakis Na. Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. Social science & medicine. 2014 aug;125(1):60–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442969.

[18] Coleman J, Katz E, Menzel H. The Diffusion of an Innovation among Physicians. Sociometry. 1957;20(4):253–270.

[19] VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Statistics in medicine. 2013 feb;32(4):591–6; discussion 597–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23341081.

[20] Lienert J, Marcum CS, Finney J, Reed-Tsochas F, Koehly L. Social influence on 5-year survival in a longitudinal chemotherapy ward co-presence network. Network Science. 2017;5(3):308–327.

[21] Hughes JJ. Paying injection drug users to educate and recruit their peers: why participant-driven interventions are an ethical public health model. Quality management in health care. 1999;7(4):4–12.

[22] Ssali S, Wagner G, Tumwine C, Nannungi A, Green H. HIV Clients as Agents for Prevention: A Social Network Solution. AIDS research and treatment. 2012;2012:815823.

[23] Cislaghi B, Denny EK, Cisse´ M, Gueye P, Shrestha B, Shrestha PN, et al. Changing Social Norms: the Importance of "Organized Diffusion" for Scaling Up Community Health Promotion and Women Empowerment Interventions. Prevention Science. 2019;p. 1–11.

[24] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. Aids. 2005;19:S67–S72.

[25] Smith R, Rossetto K, Peterson BL. A meta-analysis of disclosure of one's HIV-positive status, stigma and social support. AIDS care. 2008;20(10):1266–1275.

[26] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Social problems. 1989;36(4):416–430.

[27] Benjamin-chung J, Abedin J, Berger D, Clark A, Jimenez V, Konagaya E, et al. Spillover effects on health outcomes in low- and middle-income countries: a systematic review. International Journal of Epidemiology. 2017;(September):1–26.

[28] Neema S, Atuyambe LM, Otolok-Tanga B, Twijukye C, Kambugu A, Thayer L, et al. Using a clinic based creativity initiative to reduce HIV related stigma at the Infectious Diseases Institute, Mulago National Referral Hospital, Uganda. African health sciences. 2012;12(2):231–239.

[29] Smith Fawzi MC, Siril H, Li N, McAdam K, Ainebyona D, McAdam E, et al. Agents of Change: stepped wedge randomized controlled trial of the NAMWEZA intervention with people living with HIV and members of their social networks. BMJ Global Health. 2019;.

[30] Kaaya S, Arthur G, Siril H, Lienert J, Li N, McAdam K, et al. Leadership of PLWH as Agents of Change: comparing HIV-related risk behavior of people attending ART clinics in Dar es Salaam with members of their social networks. In: 20th International AIDS Conference; 2014..

[31] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC medical research methodology. 2006;6(1):54.

[32] Zou J, Yamanaka Y, John M, Watt M, Ostermann J, Thielman N. Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health. 2009;9(1):75.

[33] McCarthy B. Appreciative Inquiry: an alternative to behaviour management. Dementia. 2017;16(2):249–253.

[34] Siril HN, Kaaya SF, Fawzi MKS, Mtisi E, Somba M, Kilewo J, et al. CLINICAL outcomes and loss to follow-up among people living with HIV participating in the NAMWEZA intervention in Dar es Salaam, Tanzania: a prospective cohort study. AIDS research and therapy. 2017;14(1):18.

[35] Carey MP, Schroder KE. Development and psychometric evaluation of the brief HIV Knowledge Questionnaire. AIDS education and prevention. 2002;14(2):172–182.

[36] Haile BJ, Chambers JW, Garrison JL. Correlates of HIV knowledge and testing: Results of a 2003 South African survey. Journal of Black Studies. 2007;38(2):194–208.

[37] Aarø LE, Breivik K, Klepp KI, Kaaya S, Onya HE, Wubs A, et al. An HIV/AIDS knowledge scale for adolescents: item response theory analyses based on data from a study in South Africa and Tanzania. Health education research. 2011;26(2):212–224.

[38] Peltzer K, Matseke G, Mzolo T, Majaja M. Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. BMC public health. 2009;9(1):174.

[39] Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry. 2009;8(1):50-61.

[40] Easterly W. How the millennium development goals are unfair to Africa. World development. 2009;37(1):26–35.

[41] Barbarin OA, Khomo N. Indicators of economic status and social capital in South African townships: What do they reveal about the material and social conditions in families of poor children? Childhood. 1997;4(2):193–222.

[42] Smith JA, McPherson M, Smith-Lovin L. Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. American Sociological Review. 2014 Jun;79(3):432-56.

[43] Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to censor? American journal of epidemiology. 2017;187(3):623–632.

[44] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.). 2010;21(1):13.

[45] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[46] Hong, G., 2015. Causality in a social world: Moderation, mediation and spill-over. John Wiley & Sons.

[47] Forastiere L, Mealli F, Vanderweele TJ. Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs: Disentangling Bed Nets Using Bayesian Principal Stratification Identification and Estimation of Causal Mechanisms in Clustered Encouragement Designs : Disentangling Be. Journal of the American Statistical Association. 2016;111(514):510–525. Available from: http://dx.doi.org/10.1080/01621459.2015.1125788.

[48] Imai K, Jiang Z. Identification and sensitivity analysis of contagion effects in randomized placebo-controlled trials. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2019.

[49] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. SMR. 2012;40(2):211–39.

[50] De La Haye K, Robins G, Mohr P, Wilson C. Homophily and contagion as explanations for weight similarities among adolescent friends. Journal of Adolescent Health. 2011;49(4):421–427.

[51] Vanderweele TJ, Hong G, Jones SM, Brown JL. Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. Journal of the American Statistical Association. 2013;108(502):469–482. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23997375.

[52] Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013;18(2):137.

[53] Leung MP. Treatment and spillover effects under network interference. Review of Economics and Statistics. 2016;p. 1–42.

[54] Latkin CA, Davey-Rothwell MA, Knowlton AR, Alexander KA, Williams CT, Boodram B. Social network approaches to recruitment, HIV prevention, medical care, and mediation adherence. Journal of Acquired Immune Deificiency Syndromes. 2013;63:1–9.

[55] Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643–653.

[56] Berkman LF. Social support, social networks, social cohesion and health. Social work in health care. 2000 Sep 26;31(2):3-14.

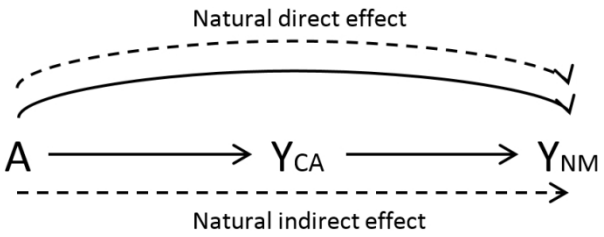
[57] Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997;25(1):95–100.

[58] Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Social science & medicine. 2015;125:192–202.

[59] He P, Eriksson F, Scheike TH, Zhang MJ. A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. Scandinavian Journal of Statistics. 2016 Mar;43(1):103-22.

11. Figure Captions

Fig 1. Schematic of natural direct effect (NDE) and natural indirect effect (NIE). The NDE indicates the increase in NMs' HIV knowledge happens without a concomitant increase in their CA's HIV knowledge. The NIE, on the other hand, indicates that the increase in NMs' HIV knowledge is mediated by their CA's HIV knowledge increasing. Solid lines indicate paths of causality between variables. Dashed lines represent the line or lines composing the effect of interest.



A: Intervention status of the CA
Y_{CA}: HIV knowledge of the CA at the interview immediately prior to the NM's follow-up interview
Y_{NM}: HIV knowledge of the NM at follow-up

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	4-5
Methods			
Study Design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	
		Case-control study—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	6
Study Size	10	Explain how the study size was arrived at	7
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6-7
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	8
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key Results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.