Social network targeting to maximise population behaviour change: a cluster randomised controlled trial

David A Kim, Alison R Hwong, Derek Stafford, D Alex Hughes, A James O'Malley, James H Fowler, Nicholas A Christakis

Summary

Background Information and behaviour can spread through interpersonal ties. By targeting influential individuals, health interventions that harness the distributive properties of social networks could be made more effective and efficient than those that do not. Our aim was to assess which targeting methods produce the greatest cascades or spillover effects and hence maximise population-level behaviour change.

Methods In this cluster randomised trial, participants were recruited from villages of the Department of Lempira, Honduras. We blocked villages on the basis of network size, socioeconomic status, and baseline rates of water purification, for delivery of two public health interventions: chlorine for water purification and multivitamins for micronutrient deficiencies. We then randomised villages, separately for each intervention, to one of three targeting methods, introducing the interventions to 5% samples composed of either: randomly selected villagers (n=9 villages for each intervention); villagers with the most social ties (n=9); or nominated friends of random villagers (n=9; the last strategy exploiting the so-called friendship paradox of social networks). Participants and data collectors were not aware of the targeting methods. Primary endpoints were the proportions of available products redeemed by the entire population under each targeting method. This trial is registered with ClinicalTrials.gov, number NCT01672580.

Findings Between Aug 4, and Aug 14, 2012, 32 villages in rural Honduras (25–541 participants each; total study population of 5773) received public health interventions. For each intervention, nine villages (each with 1–20 initial target individuals) were randomised, using a blocked design, to each of the three targeting methods. In nomination-targeted villages, 951 (74·3%) of 1280 available multivitamin tickets were redeemed compared with 940 (66·2%) of 1420 in randomly targeted villages and 744 (61·0%) of 1220 in indegree-targeted villages. All pairwise differences in redemption rates were significant (p<0·01) after correction for multiple comparisons. Targeting nominated friends increased adoption of the nutritional intervention by 12·2% compared with random targeting (95% CI 6·9–17·9). Targeting the most highly connected individuals, by contrast, produced no greater adoption of either intervention, compared with random targeting.

Interpretation Introduction of a health intervention to the nominated friends of random individuals can enhance that intervention's diffusion by exploiting intrinsic properties of human social networks. This method has the additional advantage of scalability because it can be implemented without mapping the network. Deployment of certain types of health interventions via network targeting, without increasing the number of individuals targeted or the resources used, could enhance the adoption and efficiency of those interventions, thereby improving population health.

Funding National Institutes of Health, The Bill & Melinda Gates Foundation, Star Family Foundation, and the Canadian Institutes of Health Research.

Published **Online**
May 5, 2015
http://dx.doi.org/10.1016/S0140-6736(15)60095-2

See **Online/Comment**
http://dx.doi.org/10.1016/S0140-6736(15)60503-7

Interfaculty Initiative in Health Policy, Harvard University, Cambridge, MA, USA (D A Kim BSc, A R Hwong BSc); **Yale Institute for Network Science, Yale University, New Haven, CT, USA** (D Stafford BSc); **Department of Political Science, University of California, San Diego, La Jolla, CA, USA** (D A Hughes BSc); **Dartmouth Institute of Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA** (Prof A J O'Malley PhD); **Departments of Political Science and Medical Genetics, University of California, San Diego, La Jolla, CA, USA** (Prof J H Fowler PhD); **and Departments of Medicine, Ecology and Evolutionary Biology, Biomedical Engineering, and Sociology, Yale University, New Haven, CT, USA** (Prof N A Christakis MD)

Correspondence to:
Prof Nicholas A Christakis, Yale Institute for Network Science, PO Box 208263, New Haven, CT 06520-8263, USA
nicholas.christakis@yale.edu

Introduction

Advances in understanding of the structure[1–3] and function[4,5] of social networks have opened new frontiers for interventions to improve the health of individuals and populations.[6–9] Because knowledge and behaviour can spread across interpersonal ties,[10,11] and because the networks formed by such ties tend to amplify this spread,[12–14] changes in one person's behaviour can cascade out across a social network, producing behavioural changes in other people in the population-at-large. Such cascades offer the prospect of increasing the effectiveness of public health campaigns that seek to disseminate salubrious practices, and could prove especially beneficial in low-resource settings.[15]

Deliberately fostering cascade effects requires the identification of potentially influential individuals among whom to launch an intervention. However, whom in a social network to target with the relevant knowledge or behaviour so as to maximise such diffusion is not clear. Simulation results suggest, for instance, that targeting highly connected (or high "degree") individuals in networks could enhance the population-level efficacy of prophylactic interventions.[16,17] Other research, meanwhile, suggests more complex methods for the optimal targeting of interventions.[14,18,19]

Most such methods require mapping whole social networks to identify targets. Such mapping is costly, time-consuming, and often infeasible in real-world, face-to-face situations. If network analysis is to meaningfully inform the design of policy and interventions, then simple, cost-effective procedures must be developed to identify structurally influential individuals without mapping their

entire networks. We therefore explore both a conventional measure of network centrality (so-called indegree, defined as the number of times a person is named as a social contact by other people), and an alternative strategy that does not require ascertainment of global network structure (namely, seeding a network via the friends of randomly selected individuals). The latter strategy exploits the so-called friendship paradox of human social networks: on average, the friends of randomly selected individuals are more central in the network than the individuals who named them; colloquially, "your friends have more friends than you do."[7,20] But despite its theoretical promise and its demonstrated efficacy in the early detection of outbreaks,[7] friendship nomination, to our knowledge, has never been tested experimentally as a targeting strategy for a real-world network intervention. If effective, this strategy would identify targets likely to foster cascades without mapping whole networks, a crucial condition for scalable network strategies in resource-limited settings.

We therefore conducted a randomised controlled trial of network targeting algorithms using two common but dissimilar public health interventions: chlorine for water purification and multivitamins for micronutrient deficiencies. Our aim was to assess which targeting methods produce the greatest cascades or spillover effects and hence maximise population-level behaviour change. In comparable villages, we delivered the same public health interventions to the same fraction of the population (5%), varying only the method of selecting targets. Because we are interested in the production of spillovers that change the knowledge and behaviour of untargeted individuals, we prospectively followed all members of all villages to measure their knowledge and behaviour with respect to the interventions.

Methods

Study design and participants

In this cluster randomised controlled trial, participants were recruited from villages of the Department of Lempira, Honduras. Lempira is a rural, mountainous, coffee-growing region in which geographic barriers and little transportation tend to isolate villages from one another. All villages in the area were eligible for inclusion; however, eligible individuals had to be at least 15 years old. Between 79·5% and 96·8% of all adults in each village participated in the study, with an overall participation rate of 86·7%.

We measured the entire social network of each village by asking all residents to identify spouses, siblings, and friends from a photographic census (appendix p 2). We subsequently conducted a public health needs assessment with community leaders, who identified diarrhoeal illness and nutritional deficiencies as prevalent local health concerns; we therefore selected chlorine for water purification and multivitamins for micronutrient deficiencies as the public health interventions to deploy.

See Online for appendix

We conducted a baseline survey to assess knowledge, attitudes, and behaviours surrounding water purification and nutrition, and followed villagers (whether targeted or not) for their knowledge and behaviours (appendix p 2).

The study was approved by the Institutional Review Board of Harvard Medical School. All participants provided informed consent.

Randomisation and masking

In the design of this study, we used design-of-experiments principles, including randomisation, blocking, and the ideas of orthogonalisation and balanced treatment assignments (applied with respect to the handling of the two interventions; appendix p 3) to obtain precise and unbiased estimates of the effects of three network targeting strategies (random, highest indegree, and friendship nomination) on the diffusion and uptake of two health interventions (multivitamins for micronutrient deficiencies and chlorine for water purification). We randomised villages to targeting mechanisms to ensure that the distribution of potential (observed and unobserved) confounding variables would be the same (in expectation) across the villages.

To better isolate the effect of targeting method from potential village-level influences on product adoption, which in 32 villages could be unevenly distributed across targeting methods by chance alone, we divided the 32 villages into eight blocks on the basis of network size, mean socioeconomic status, and baseline rates of water purification (appendix pp 3–4, 8–10). We used the results of a factor analysis to form an aggregate score that explained most of the variance in the three blocking variables. We then found the assignment of villages to blocks that minimised the ratio of within-block to between-block variance in the distribution of the composite score (appendix p 3).

After blocking, we randomly assigned each village in each block to targeting methods for multivitamins and chlorine (indegree targeting, nominated friends targeting, random targeting, or no intervention), using a fractional-factorial design (appendix pp 3, 9). That is, for each of the two interventions separately, nine of the 32 villages were targeted randomly, nine by highest indegree, and nine by the nominated friends technique. Six villages received only one intervention and two villages received neither intervention. For each village receiving both multivitamin and chlorine interventions, we used a different targeting method for each intervention. Participants and data collectors were not aware of the targeting methods.

Procedures

In indegree-targeted villages, we targeted the 5% of villagers named as a contact most often by others in their village. In the nomination-targeted villages, we targeted a 5% sample of villagers composed of one randomly chosen friend nominated by each member of a 5% random sample of villagers. In the randomly targeted

villages, we targeted a random 5% sample of villagers (see appendix p 4 for details of targeting algorithms). Figure 1 shows the targeting of each intervention for a representative block of villages. In the weeks following the introduction of the interventions, we tracked the diffusion of products and knowledge among all villagers, targeted and untargeted.

During the course of 1 day for each village, we delivered to each targeted individual an intervention consisting of a health product, instructions for use, and an educational

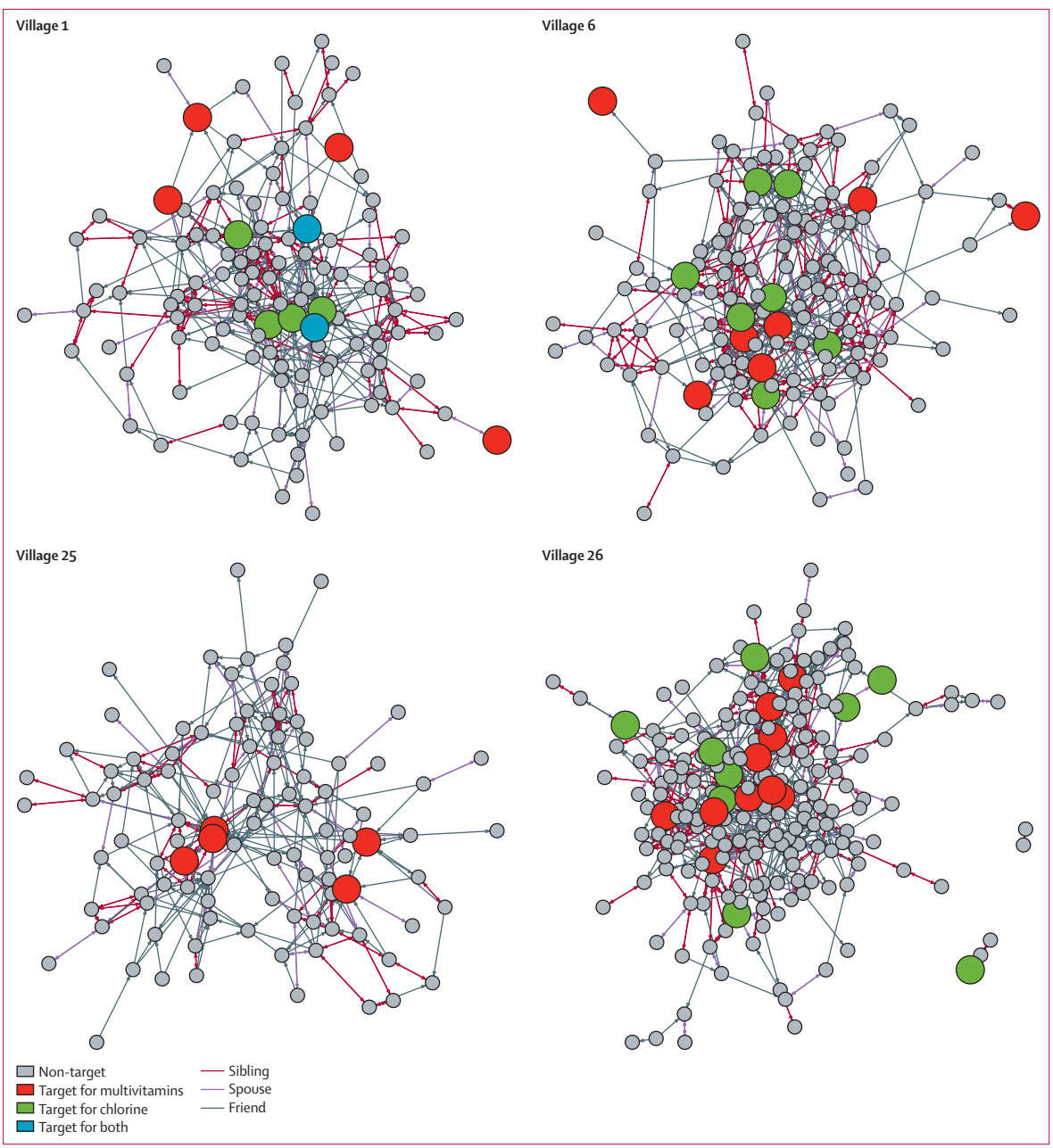

Figure 1: **Targeting of interventions for one block of villages**
Within each block of four villages, we randomly assigned each village to a targeting method (indegree, nominated, random, or none). We did this for each of the two interventions in a fractional factorial design (appendix p 9). In the indegree-targeted villages, we targeted the 5% of villagers with highest indegree. In the nomination-targeted villages, we targeted one randomly chosen friend nominated by each member of a 5% random sample of villagers. In the randomly targeted villages, we targeted a 5% random sample of villagers. In the villages receiving both interventions (chlorine and vitamins), a small proportion of villagers were drawn, by chance, as targets for both interventions (eg, two blue nodes in village 1). In the block shown, village 1 received multivitamins (red nodes) by random targeting and chlorine (green nodes) by nomination targeting. Village 6 received multivitamins by nomination targeting and chlorine by indegree targeting. Village 25 received multivitamins only, by indegree targeting. Village 26, finally, received multivitamins by indegree targeting and chlorine by random targeting. Some visible groups of siblings are not fully interconnected given a deliberate feature of the "name generator" used to map the network (appendix p 2).

component (appendix p 6). We also gave targeted individuals supplementary information about the interventions that was not generally known at baseline or circulated by other means, and asked them to relay this information to others, allowing us to track the diffusion of knowledge as well as of product adoption by the study's completion. Multivitamin targets received 60 adult multivitamins (see appendix p 2 for formulation). Chlorine targets received a 250 mL plastic bottle of sodium hypochlorite with a medicine dropper. A small proportion of villagers were drawn, by chance, as targets for both products, and so received both (appendix p 4).

Each targeted villager also received four tickets to distribute to friends or family outside the household but within the village, who could in turn redeem the ticket for the same product at a local store for a nominal fee to the shopkeeper. Targets were asked to instruct the villagers to whom they gave tickets about the correct usage and benefits of the product. Each of these initial (first-wave) ticket redeemers received, with their product, a packet of four additional (second-wave) tickets for distribution to additional villagers. Every ticket was uniquely identified and was signed, dated, and checked by a participating shopkeeper against a list of eligible study participants upon redemption, which enabled us to track the diffusion of products through the village networks with time.

We used ticket redemption for multivitamins and chlorine as our primary measure of behaviour because ticket redemption was the most accurately and comprehensively recorded measure of product uptake (ie, we know the identity without exception of every individual who redeemed a ticket, and the exact date on which she did so), and because it allowed us to trace with the greatest social and temporal resolution the rate and extent of product diffusion through the village networks, without relying on participants' recollection or self-report.

We supplemented this objective behavioural measure (ticket redemption) with self-reports of knowledge and practice as well, conducting an extensive follow-up survey in all villages, 4–6 weeks after the interventions, in which we asked villagers (whether or not they had redeemed a ticket) about their use of the products, their attitudes concerning the products' utility and effectiveness, and a series of factual questions about their correct usage and benefits (from which the knowledge scores, our secondary outcomes, were derived, as detailed in appendix p 6).

After completion of the entire trial, we donated additional multivitamins and other supplies to all villages in the study.

Outcomes

The primary outcomes were the proportions of available products redeemed by the entire population under each targeting method. We evaluated the basic diffusion of the products by calculating, for each day after the initial delivery of interventions to the targeted villagers, the proportion of available tickets redeemed for each product in each group of villages (ie, indegree, nomination, and random targeting). Each of these tickets was redeemed by a study participant who had received a ticket from a targeted villager or from a first-wave ticket redeemer. Both waves of ticket redemption are pooled in our population-level analyses.

The secondary outcomes were the proportions of villagers under each targeting method attaining high knowledge scores. To assess knowledge transmission, we formed composite 0–10 scores for each intervention, using the first component from a principal components analysis of the knowledge and usage questions from the follow-up survey (appendix p 6). We then calculated, for each intervention, the proportion of respondents under each targeting method (indegree, nomination, and random) achieving scores in the top quartile.

Statistical analysis

We evaluated the effect of targeting method at both the population and the individual level. At the population level, we used χ^2 tests to assess differences in the proportions of tickets redeemed or high scores attained across the three targeting methods, for each intervention.

To ensure that our aggregate results were not driven by variation in individual-level or village-level characteristics other than targeting method, we also estimated the effect of targeting method on the hazard of individual tickets being redeemed, using mixed-effects Cox models to control for both individual-level and village-level characteristics, and controlling for possible interference between the two interventions in villages receiving both (appendix p 5). We also estimated the effect of targeting method on knowledge transmission with multilevel logit models of high knowledge score attainment, again controlling for both individual-level and village-level characteristics (appendix pp 6–7).

We account for the presence of simultaneous interventions in certain villages in both the design and the analysis of the study. At the design stage, we use multiple instances of all targeting methods for both interventions, and permute the combinations across villages so as to best mitigate potential interference between interventions, and to allow for simple interference tests to be performed (appendix p 3). In our mixed-effect models of ticket redemption (appendix pp 11–12), we control for interference between the two interventions by including a ticket redeemer's redemption of a ticket for the opposite intervention as a time-varying covariate (appendix p 5). For models of high knowledge score attainment (appendix p 13), we estimate the effects of the targeting methods of both products on the attainment of a high knowledge score for either product.

We calculated confidence intervals for the ratios of binomial parameters (such as ticket redemption rates) using a skewness-corrected likelihood score-based method,[21] and for the ratios of mean target group

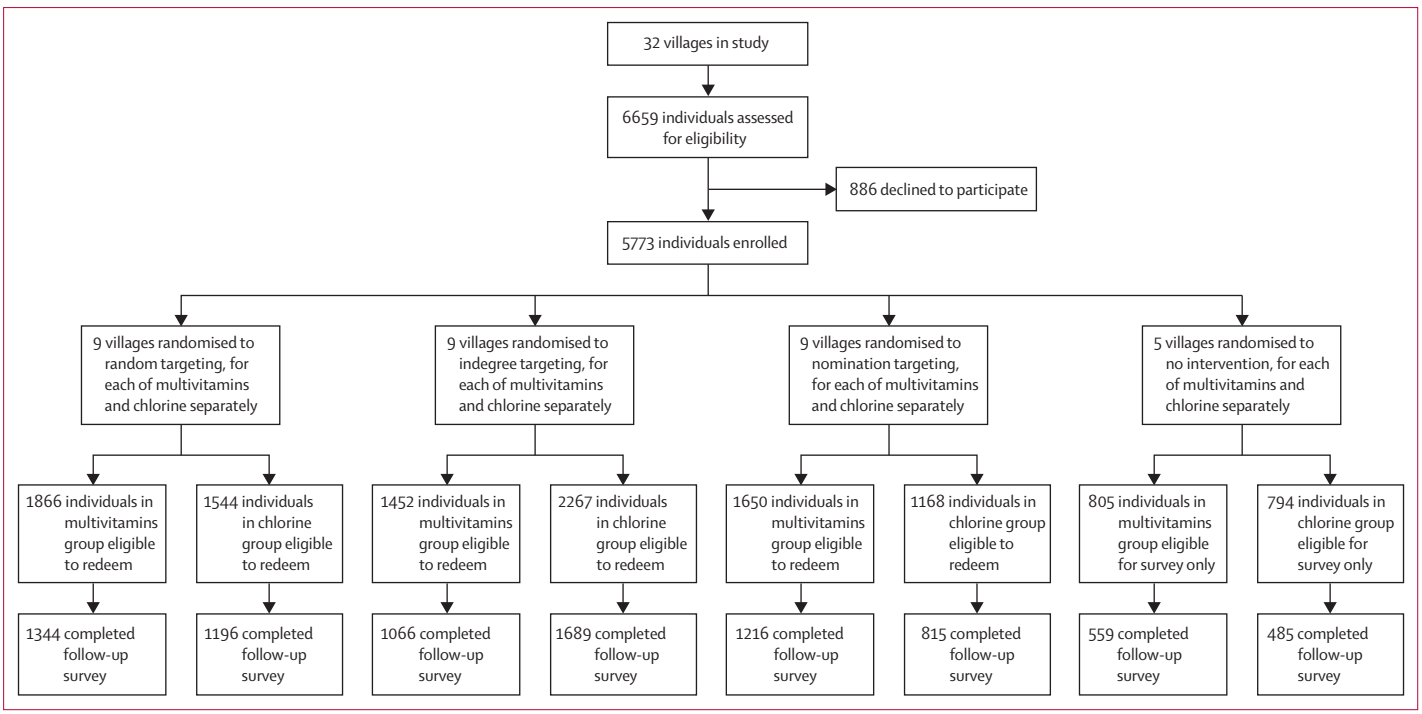

Figure 2: **Trial profile**

Primary endpoint was the proportion of available products (multivitamins or chlorine) redeemed by the entire population under each targeting method, where the number of available products in each village was proportional to the number of villagers for whom network data was available (see appendix p 4 for details). Thus, the primary treatments (targeting method: random, indegree, or nomination) and the primary endpoints (proportion of available products redeemed under each targeting method) are interpreted at the population level. As such, no participants dropped out or were lost to follow-up with respect to the primary outcome, because the primary outcome was defined as the proportion of available products redeemed under each targeting method, and because the identity of the ticket redeemer and the date of redemption were recorded upon the ticket's redemption (see appendix pp 4–5 for details). Completion of the follow-up survey pertains only to secondary outcomes (eg, knowledge scores). For primary outcome (ticket redemption), the disposition of all available tickets is known, and no data on the diffusion of products (the primary outcome) were lost, with the exception of one village in which the ticket-redemption process was imperfectly documented: inclusion or exclusion of this village's data do not affect the main results (see appendix p 7 for details).

properties using a method adjusted for heteroscedasticity.[22] All analyses were performed in *R* (version 3.1).

This trial is registered with ClinicalTrials.gov, number NCT01672580, and we did not deviate from our original analytic plan.

Role of the funding source

The trial was an investigator-initiated study supported by grants from the National Institutes of Health, the Bill & Melinda Gates Foundation, the Star Family Foundation, and the Canadian Institutes of Health Research. None of the funding sources had any role in the design, conduct, or analysis of the study, the writing of the report, or the decision to submit it for publication. All authors had full access to all the data in the study. NAC had final responsibility for the decision to submit for publication.

Results

Between Aug 4, 2012, and Aug 14, 2012, 32 villages in rural Honduras (25–541 participants each; total study population of 5773) received public health interventions (figure 2). 2944 (51%) of the 5773 participants were women. The mean age was 35 years (SD 14), 1160 (20·1%) respondents were unmarried, and respondents had a mean of 4·2 years (SD 2·4) of formal schooling.

For each intervention, nine villages (each with 1–20 initial target individuals) were randomised to each of the three targeting methods (figure 2).

Consistent with network theory, the targeting methods succeeded in identifying different segments of the village networks in the three treatment groups of the study. Targets in indegree-targeted villages had 2·2 times (95% CI 2·0–2·5) the mean indegree of randomly chosen targets, and 1·6 times (1·4–1·9) the mean indegree of targets chosen by the nomination method. These nominated targets, meanwhile, had 1·4 times (95% CI 1·2–1·6) the mean indegree of random targets.

First, we evaluated ticket redemption rates (representing product adoption) and knowledge scores at follow-up (representing knowledge diffusion) for the multivitamin intervention (figure 3, left panel). In nomination-targeted villages, 951 (74·3%) of 1280 available multivitamin tickets were redeemed compared with 940 (66·2%) of 1420 in randomly targeted villages, and 744 (61·0%) of 1220 in indegree-targeted villages. All pairwise differences in redemption rates were significant ($p<0·01$) after correction for multiple comparisons (appendix p 5). Hence, targeting nominated friends increased population-level adoption of a nutritional intervention by 12·2% (95% CI 6·9–17·9), an 8·1 % point increase in product

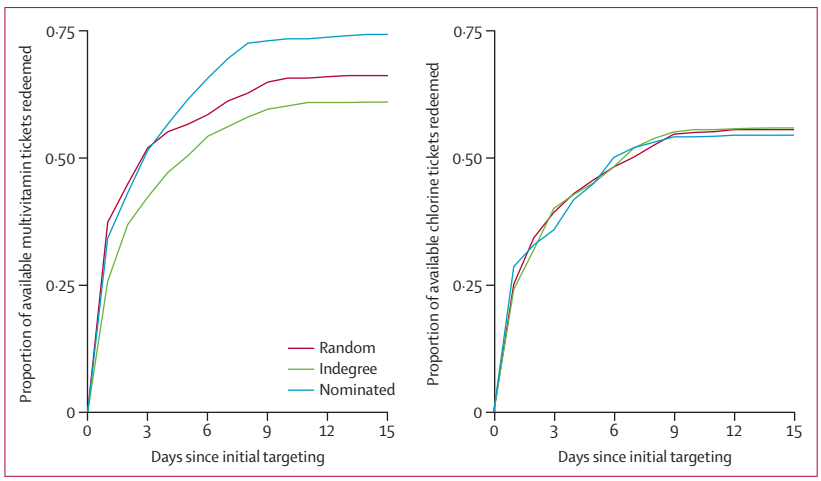

Figure 3: **Diffusion of interventions**
The left panel shows the pooled proportion of available multivitamin tickets redeemed by day after initial targeting, by treatment group. The right panel shows the equivalent measure for the chlorine intervention.

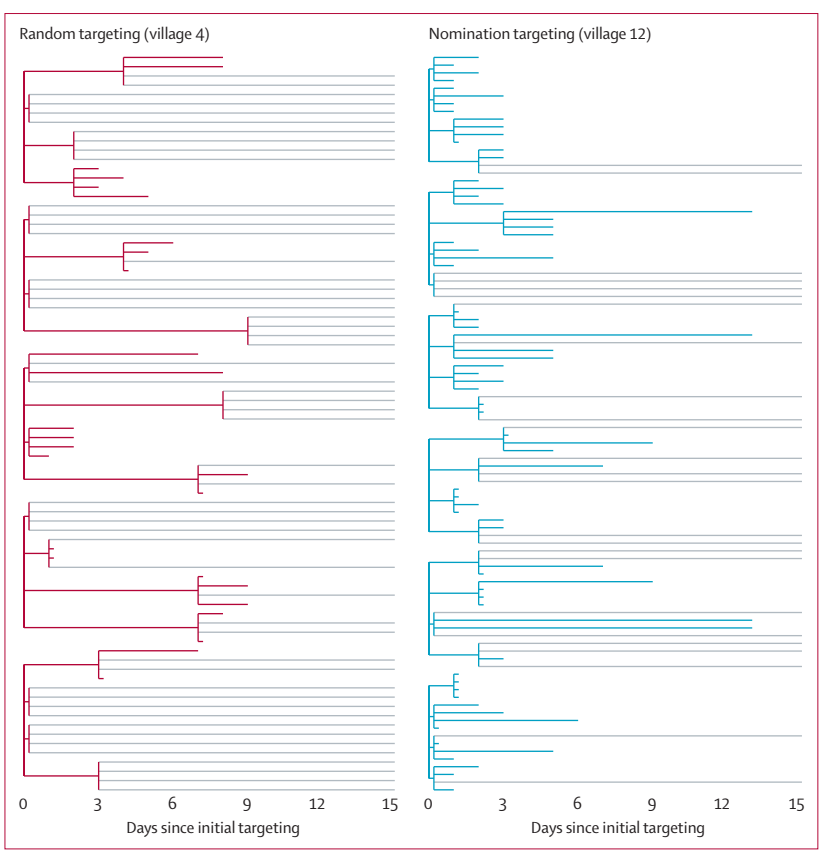

Figure 4: **Multivitamin diffusion through two representative villages**
In all villages receiving a given intervention, 5% of the adult population was initially targeted with the intervention. The left panel shows the diffusion of multivitamins in a village whose five initial targets were selected at random. The right panel represents a village of comparable size whose six initial targets were the nominated friends of random villagers. Every targeted villager received four tickets to distribute to contacts outside the household, who could in turn redeem the ticket for the same product and for four additional tickets to give to others. The coloured lines represent completed ticket-redemption paths: shorter lines represent tickets redeemed more quickly. Light grey lines represent available tickets that were not redeemed. Compared with the randomly targeted village, multivitamins in the nomination-targeted village diffused more rapidly (the average time-to-redemption of first-wave tickets is substantially shorter) and more completely (a greater proportion of available tickets was ultimately redeemed).

adoption. Figure 4 shows the effect of targeting method on product diffusion cascades for two representative villages: compared with random targeting, nomination targeting produced both more rapid and more thorough adoption of the multivitamin intervention.

These results concord with our multilevel Cox models of ticket redemption, in which we estimate a hazard ratio of 1·65 (95% CI 1·10–2·47) for redemption of first-wave tickets under nomination targeting, compared to random targeting, and controlling for individual-target-level and village-level characteristics (appendix p 11). For the redemption of second-wave tickets, our estimated hazard ratio of 1·15 under nomination targeting compared to random targeting does not reach statistical significance (p=0·54); a larger sample may have allowed us to detect effects out to two degrees of separation from the initial targets.

With respect to knowledge, 30·8% of untargeted ticket recipients attained high knowledge scores in nomination-targeted villages, compared with 27·6% in both indegree and randomly targeted villages. Although the differences in aggregate proportions are not statistically significant (p>0·05), a multilevel logit model controlling for both individual-level and village-level characteristics (appendix p 13) reveals a positive and statistically significant effect of nomination targeting (compared with random targeting) on the attainment of high multivitamin knowledge scores (OR 1·66 [95% CI 1·02–2·70]).

Conditional on ticket redemption, no significant variation was noted in self-reported product use or belief in the products' effectiveness by targeting method (random, indegree, or friend nomination). Self-reported continued product use among confirmed ticket redeemers was uniformly high (>90%). These results suggest that ticket redemption (in which we did observe significant variation by targeting method, both at the population level, as well as in mixed-effects models accounting for individual-level and village-level co-variates and for possible interference between interventions) is indeed a valid omnibus measure of product adoption and continued use, while the knowledge scores we report provide additional data about the differential spread of health information not predicted by product use alone.

We accounted for potential interference between the two interventions at both the village and the individual level. All such robustness tests support a positive effect of nomination targeting on the uptake of the multivitamin intervention (appendix pp 3, 5–6, and 11). The mean village-level multivitamin ticket redemption rate is statistically indistinguishable between those villages that received both multivitamin and chlorine interventions (on average, 73·1% of these villages' available multivitamin tickets were redeemed), and those villages that received the multivitamin intervention alone (on average, 73·8% of these villages' available multivitamin tickets were redeemed); note that these are

village-level redemption rates, as distinct from the pooled proportions depicted in figure 3. Likewise, controlling for possible interference in our mixed-effect models does not alter the trend we observe at the village level (appendix p 11).

Finally, we undertook corresponding analyses for the chlorine intervention (figure 3, right panel). For this intervention, final ticket redemption rates were statistically indistinguishable (p>0·05) across targeting methods: 55·6% for random targeting, 55·9% for indegree, and 54·5% for nomination. Although we observe the same trend in knowledge scores as for multivitamins (38·4% high scores in nomination-targeted villages, compared with 36·0% in indegree targeted villages and 35·0% in randomly targeted villages), the aggregate differences are not statistically significant. Corresponding individual-level models likewise showed no significant effect of targeting method on ticket redemption or knowledge scores in the case of the chlorine intervention (p>0·05; appendix pp 12–13).

Discussion

In a randomised controlled trial in 32 villages, we evaluated network-based approaches to maximise population-level behaviour change. Our results show no evidence that health interventions benefit from targeting only the most highly connected individuals. However, a second technique, in which we selected targets by exploiting the friendship paradox, produced significantly larger cascades of product adoption and health knowledge than either random or indegree targeting (panel). For our nutritional intervention, targeting nominated friends increased population-level behaviour change by 12·2% compared with random targeting; it was also associated with enhanced health knowledge among untargeted individuals.

Of our two network targeting methods, indegree targeting requires the expenditure of substantial resources to map the whole network, because everyone in the population must be asked to whom they are connected to identify the individuals receiving the most nominations. It is therefore fortunate that the friendship nomination technique, which exploits the properties of human social networks without requiring that the entire network be mapped, produced the greatest behavioural cascades. Compared with methods requiring whole-network (sociocentric) data, targeting the nominated friends of a randomly selected group can furnish the benefits of network targeting in a more scalable and less resource-intensive fashion. Indeed, because high-degree individuals in human social networks tend to be friends with one another,[3,23] targeting nominated friends might outperform targeting the highest-degree individuals if the latter strategy produces redundant clustering among targets, resulting in an echo chamber of influence that fails to reach more dispersed or peripheral parts of the network.

Panel: Research in context

Systematic review

We searched Google Scholar for articles published in English between 1990, and 2012, with the search terms "social network" OR "social networks" OR "network" AND "intervention" OR "trial". From our review of the available literature, we concluded that there have been numerous observational studies of contagion in social networks, but very few efforts to exploit network phenomena to maximise the diffusion of desirable knowledge or behaviour with respect to health. The few trials conducted to this end have typically involved either small face-to-face populations or large online networks. We noted evidence that a randomised, prospective, large-scale trial of network targeting in a face-to-face population could contribute to a better understanding of social networks and health, with implications for the design of public health interventions in the developed and developing world.

Interpretation

In 32 villages, we evaluated network-based approaches to maximise population-level behaviour change. For a nutritional intervention, targeting the friends of randomly selected individuals produced significantly greater population-level adoption and health knowledge than targeting either random or highly connected individuals. This method has the advantage of scalability, because it can be implemented in the field without mapping the network. Our findings suggest that network targeting can be used to efficiently increase the adoption of certain types of public health interventions. Further trials will be needed to characterise the targeting methods best suited to different classes of interventions.

This could be especially likely in networks with meaningful community structure, in which there are subgroups of interconnected individuals, each with their own locally influential nodes: in such networks, indegree targeting risks selecting nodes only from the most densely connected subgroup. Our experimental results concord with recent simulations in which indegree-based targeting produces lower adoption rates than even random targeting.[19] By combining virtues of random targeting (namely, the selection of targets dispersed through the network) and highest-indegree targeting (the selection of highly connected and potentially influential individuals), the nomination method may prove more robust than either across a range of real-world network structures.

Social networks amplify the information and behaviours with which they are seeded,[10,11,24] but the nature of this diffusion depends on the innovation being transmitted.[6,25,26] The complexity of understanding and implementing a new practice, the visibility of its results, and its perceived advantage over existing methods can all affect adoption patterns.[26] Simple information (regarding, for instance, the availability of subsidised

multivitamins) can spread by simple contagion, requiring only one contact for transmission between two individuals. Deeper behavioural changes, by contrast, might require reinforcement from multiple social contacts (complex contagion),[6,25] perhaps because they require great motivation (as in the case of behaviours like smoking cessation), or because they require changes in longstanding practices and beliefs (as in the case of water chlorination in rural Honduras).

We selected for our study two interventions with very different social and behavioural implications: chlorine is widely available and affordable, even by rural Honduran standards, but its main use is for washing clothing. Chlorinating water is a fairly complex, multistep process. Thus, the chlorine intervention demanded the acquisition of new knowledge and a change in the use of a familiar product. By contrast, multivitamins in rural Honduras are easy to use and widely viewed as beneficial, but are relatively more difficult and costly to obtain. The multivitamin intervention, then, demanded substantially less behavioural and ideational change. Consistent with the theory of complex contagion,[6,25] these fundamental differences between our two interventions might account in part for the robust success of nomination targeting in increasing the adoption of multivitamins, but not of chlorine for water purification, for which an altogether different targeting method might have been effective.

Diarrhoeal illness and malnutrition account for a sizeable burden of disease in rural Honduras,[27,28] which is why we assessed water purification and multivitamins in our setting. But countless other health interventions in the developed and developing world stand to benefit from network targeting, including immunisation,[16,29] anti-malarial bednets, maternal health care, safe sex practices, smoking cessation,[30] substance abuse prevention,[12] helmet and seatbelt use, and many more. The extraordinary variety of health states and behaviours noted to spread through social networks[4,5,11] probably means that no one targeting method will prove optimum for all interventions. Simulations[16,17,19,29] and observational studies[14] suggest strategies to be tested in experiments such as ours, the results of which can in turn inform the design of more accurate models of behavioural cascades. Future research will continue to characterise the targeting methods best suited to different types of interventions, and, although no universally optimum solution is likely to be found, general principles, such as the efficiency of nomination-based targeting for the diffusion of certain types of phenomena, are likely to emerge.

Our study has limitations. First, we used ticket redemption and associated survey responses as our primary measures of behaviour, and did not record the use of the products in people's homes. Second, the villages in our study were isolated from one another, which, although ideal for a controlled experiment, could yield different patterns of diffusion than would more intermingled populations. Third, longer-term follow-up in the villages would allow us to determine whether the knowledge and behavioural changes associated with a given targeting method persist with time. Although fadeout is endemic to certain educational interventions, this effect could vary according to network structure and the targeting method of the intervention.[31]

Network targeting could prove most useful in settings in which limited resources or infrastructure render broadcast interventions infeasible, but even when networks are more fully characterised (eg, with big data techniques[32]), methods to efficiently identify structurally influential individuals, as well as the people around them who are likely to be influenced, could prove useful in the design of more cost-effective campaigns. In this way, deploying health interventions via network targeting, without increasing the number of people targeted or the expenses incurred, could enhance the spread and adoption of those interventions, and thereby improve population health.

Contributors

DAK, ARH, DS, DAH, AJO'M, JHF, and NAC designed the study. DAK, ARH, DS, and DAH collected the data. DAK, AJO'M, JHF, and NAC analysed the data. DAK, AJO'M, JHF, and NAC interpreted the data. DAK, ARH, and NAC did the literature search. DAK, JHF, and NAC drew the figures and wrote the report.

Declaration of interests

NAC reports grants from the National Institutes of Health (P-01 AG031093, P-30 AG034420), from the Bill & Melinda Gates Foundation, and support from the Star Family Foundation. DAK was supported by the Canadian Institutes of Health Research and ARH was partially supported by NIH F30 AG046978. All other authors declare no competing interests.

Acknowledgments

We thank The Clorox Company and Tishcon Corporation for their donations of supplies used in the study in Honduras. We thank Laurie Meneades, Peter Treut, Jen Zhu, Michael DeWit, Bret Abel, Tom Keegan, and our local community health workers and leaders in Honduras—Bany Carballo, Mirna Castellanos, Orlin Chavez, Martina Hernandez, Marco Antonio Miranda, and Dilcia Ponce—for the expert assistance required to collect and assemble the dataset.

References

1 Newman MEJ, Barabási A-L, Watts DJ. The structure and dynamics of networks. Princeton: Princeton University Press, 2006.
2 Barabási A-L, Albert R. Emergence of scaling in random networks. *Science* 1999; **286:** 509–12.
3 Apicella CL, Marlowe FW, Fowler JH, Christakis NA. Social networks and cooperation in hunter-gatherers. *Nature* 2012; **481:** 497–501.
4 Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007; **357:** 370–79.
5 Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008; **358:** 2249–58.
6 Centola D. The spread of behavior in an online social network experiment. *Science* 2010; **329:** 1194–97.
7 Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. *PLoS One* 2010; **5:** e12948.
8 Valente TW. Network interventions. *Science* 2012; **337:** 49–53.
9 Gardy JL, Johnston JC, Sui SJH, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; **364:** 730–39.
10 Christakis NA, Fowler JH. Connected: the Surprising Power of Social Networks and How They Shape Our Lives. New York: Little, Brown & Company, 2009.
11 Fowler JH, Christakis NA. Cooperative behavior cascades in human social networks. *Proc Natl Acad Sci USA* 2010; **107:** 5334–38.

12 Valente TW, Ritt-Olson A, Stacy A, Unger JB, Okamoto J, Sussman S. Peer acceleration: effects of a social network tailored substance abuse prevention program among high-risk adolescents. *Addiction* 2007; **102:** 1804–15.

13 Rand DG, Arbesman S, Christakis NA. Dynamic social networks promote cooperation in experiments with humans. *Proc Natl Acad Sci USA* 2011; **108:** 19193–98.

14 Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO. The diffusion of microfinance. *Science* 2013; **341:** 1236498.

15 Merzel C, D'Afflitti J. Reconsidering community-based health promotion: promise, performance, and potential. *Am J Public Health* 2003; **93:** 557–74.

16 Pastor-Satorras R, Vespignani A. Immunization of complex networks. *Phys Rev E* 2002; **65:** 036104.

17 Bahr DB, Browning RC, Wyatt HR, Hill JO. Exploiting social networks to mitigate the obesity epidemic. *Obesity* 2009; **17:** 723–28.

18 Aral S, Muchnik LEV, Sundararajan A. Engineering social contagions: optimal network seeding in the presence of homophily. *Network Sci* 2013; **1:** 1–29.

19 Cho Y, Hwang J, Lee D. Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach. *Technol Forecast Soc* 2012; **79:** 97–106.

20 Feld SL. Why your friends have more friends than you do. *Am J Sociol* 1991; **96:** 1464–77.

21 Gart JJ, Nam J-M. Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics* 1988; **44:** 323–38.

22 Hasler M, Vonk R, Hothorn LA. Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Stat Med* 2008; **27:** 490–503.

23 Newman MEJ. Assortative mixing in networks. *Phys Rev Lett* 2002; **89:** 208701.

24 Bond RM, Fariss CJ, Jones JJ, et al. A 61-million-person experiment in social influence and political mobilization. *Nature* 2012; **489:** 295–98.

25 Centola D, Macy MW. Complex contagions and the weakness of long ties. *Am J Sociol* 2007; **113:** 702–34.

26 Rogers E. Diffusion of innovations. New York: Free Press, 2003.

27 World Health Organization. Country profile of environmental burden of disease: Honduras. http://www.who.int/quantifying_ehimpacts/national/countryprofile/honduras.pdf (accessed Sept 8, 2013).

28 Solórzano GJO, Molina IB, Turcios-Ruiz RM, et al. Burden of diarrhea among children in Honduras, 2000–2004: estimates of the role of rotavirus. *Rev Panam Salud Publica* 2006; **20:** 377–84.

29 Cohen R, Havlin S, ben-Avraham D. Efficient immunization strategies for computer networks and populations. *Phys Rev Lett* 2003; **91:** 247901.

30 Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. *Am J Public Health* 2010; **100:** 1282–89.

31 Cascio EU, Staiger DO. Knowledge, tests, and fadeout in educational interventions. National Bureau of Economic Research Working Paper Series, 2012; No. 18038. http://www.nber.org/papers/w18038 (accessed Sept 1, 2013).

32 Lazer D, Pentland A, Adamic L, et al. Computational social science. *Science* 2009; **323:** 721–23.

THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Kim DA, Hwong AR, Stafford D, et al. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet* 2015; published online May 5. http://dx.doi.org/10.1016/S0140-6736(15)60095-2.

WEB EXTRA MATERIAL

Social network targeting to maximise population behaviour change: a cluster randomised controlled trial

David A. Kim, M.D. Candidate

Alison R. Hwong, M.D. Candidate

Derek Stafford, Ph.D. Candidate

D. Alex Hughes, Ph.D. Candidate

A. James O'Malley, Ph.D.

James H. Fowler, Ph.D.

Nicholas A. Christakis, M.D.

Name Generators Used to Identify Social Contacts

We created a photographic census of all adults in each village and used it, with survey software we developed (publicly available in 2015), to map the social networks. Each adult was asked to identify their friends, siblings, and spouse, using the name generator questions below, and to confirm the tie by viewing a photo of the person they named. The name generators we used were:

1. Who are your brothers and sisters that you are friends with?
2. Who are your best friends that are not your brothers and sisters?
3. Who are you married to, or whom are you living with as a husband or wife?

Due to the (deliberate) phrasing with which we elicited sibling ties, not all sets of siblings are fully inter-connected (either in the graphs shown in Figure 1 or in the statistical analyses performed).

Baseline Survey Results

Prior to our intervention, we surveyed all participants regarding their practices with respect to water purification and multivitamin use (see Table S1). The mean village-level rate of water purification at baseline was 0·41 (SD=0·26). The mean village-level rate of daily multivitamin use at baseline was 0·11 (SD=0·05). Due to the unreliability of income and educational data in the region, we also asked six community health workers familiar with the villages to independently rate the socioeconomic status of each village on a 1-10 scale, and we averaged their ratings to create a score for each village; we used this average in the village-level blocking procedure described below. The resulting mean socioeconomic status (SES) rating was 4·4 (SD=2·1). Finally, the mean number of villagers with network data in each village was 146 (SD=108).

Given the limited range of multivitamin use at baseline, we chose to use water purification practices, village SES, and network size to block the villages for treatment assignment.

Choice of Interventions

We chose to study the diffusion of both multivitamins for micronutrient deficiencies and chlorine for water purification because:

1. Diarrheal illness and malnutrition account for a sizeable burden of disease in rural Honduras;
2. We conducted a needs assessment in advance of our trial in which community leaders identified multivitamins for micronutrient deficiencies and chlorine for water purification as useful to their communities, and,
3. We anticipated, based on our pre-intervention assessments and on relevant sociological theory (e.g., on simple vs. complex contagion),[1,2] that the two interventions might diffuse differently owing to their different behavioural and normative demands, thus allowing us to test for differential effects of network targeting methods on two types of interventions, rather than one.

We account for the presence of simultaneous interventions in certain villages in both the design and analysis of the study, as described in several sections below (*Blocking of Villages and Randomisation, Ticket Redemption Models, Models of Knowledge Diffusion)*.

Multivitamin Composition

The multivitamin tablets used in the study (as well as the additional multivitamins donated to all villages after the study's completion) were formulated by Tishcon Corporation as follows, and distributed in 60-count bottles:

Vitamin A 5000IU

Vitamin C 120mg
Vitamin D_3 400IU
Vitamin E 10IU
Thiamin (B_1) 1.5mg
Riboflavin (B_2) 1.7mg
Niacin 20mg
Vitamin B_6 2mg
Folic Acid 400mcg
Vitamin B_{12} 50mcg
Biotin 300mcg
Pantothenic Acid 5mg
Calcium 250mg
Iron 18mg

Blocking of Villages and Randomisation

We used the results from a factor analysis with varimax rotation to form a composite score that explained most of the variance in network size, village-level socioeconomic status, and baseline rates of water purification. The composite score can be considered a weighted average of the three constituent variables with weights reflecting the extent to which each variable accounts for variation in the composite score.

We then assigned each of the 32 villages to eight blocks that minimised the ratio of within-block to between-block variance in the distribution of the composite score. After blocking, we randomised each village to one of four targeting methods (random, indegree, nominated, or none) for the multivitamin and chlorine interventions according to the overlapping fractional-factorial designs outlined below and illustrated in Table S2. By randomising villages to targeting strategies within each block, the between-block component of variation is removed from the error term, yielding more precise results. Even under our multilevel analytic models with village random effects, the randomised block design is advantageous, since adjusting for block (as we do in the models reported in Tables S4-S6) reduces the between-village variance component.

The best way to guard against possible interference of one intervention (multivitamins or chlorine) on the other would be to perform a factorial design in which at least one village (and ideally more) received each combination of targeting methods for the two interventions (random targeting, indegree targeting, nomination targeting, or no intervention). Because we had two interventions (multivitamins and chlorine), each with three different non-null targeting mechanisms, there are $2^6 = 64$ possible arrangements of targeting mechanisms. With 32 villages, we could use only 32 of these; hence, our use of the term "fractional factorial." When including null levels of interventions (to enable estimation of interference effects or interactions of the two products), there are $2^8 = 256$ possible levels. Furthermore, we wanted to prioritize the non-null interventions as well as block on the known, observed village level-factors we thought would be related to outcomes. Therefore, as a compromise, we assigned null treatment to one village for *one* of the interventions in six blocks, and, in two additional blocks, we assigned one village to null treatment for *both* interventions (Table S2). The last two villages (i.e., those receiving no interventions) provide a potential baseline against which to evaluate the effect of a single intervention in the other six villages. The six villages receiving only one intervention, meanwhile, were assigned to targeting methods (random, indegree or nomination) at the same rates (i.e., 1/3 each) as villages receiving both interventions, and thus yield information about treatment effects that is free of any potential interference between interventions, allowing a straightforward village-level interference test to be performed.

For the multivitamin (MVI) intervention (for which we report positive effects of nomination targeting on ticket redemption and knowledge scores), the mean village-level MVI ticket redemption rate is statistically indistinguishable between those villages that received both MVI and chlorine interventions (on average, 73.1% of these villages' available MVI tickets were redeemed), and those villages that received the MVI intervention alone (on average, 73.8% of these villages' available MVI tickets were redeemed; note that these are *village*-level redemption rates, as distinct from the pooled proportions depicted in manuscript Figure 3).

In our individual-level models, we account for potential interference in much more detail, as described below.

In Table S2, Villages 1 to 4 refer to different villages across blocks but the same villages across interventions (multivitamins and chlorine). Thus, for each of the two interventions separately, nine of the 32 villages were targeted randomly, nine by highest indegree, and nine by the nominated friends technique. Six villages received only one intervention, and two villages received neither intervention.

Selection of 5% Target Groups

Table S3 shows each village's assignment to one of the eight blocks, and to one of the four targeting methods for each intervention.

In each village, 5% of the villagers for whom network data were available were selected as targets. Based on past experience, we estimated *a priori* that a 5% targeting rate provided the right balance between sufficient exposure of the villages to the tickets (i.e., the novel stimulus) on the one hand, and avoidance of immediate saturation on the other. And, empirically, our 5% targeting rate produced neither failure of the tickets to diffuse nor complete saturation in any of the study villages, resulting instead in considerable temporal and between-village variation in ticket redemption (Figure 3, main text). The observed variation allowed us to evaluate the differential effects of the three targeting methods, which we do both at the village-level (as described in the *Results* section of the main text) as well as in mixed-effects Cox models controlling for both individual-level and village-level characteristics, and testing for possible interference between the two interventions in villages receiving both (models reported in Tables S4 and S5).

In all cases, targets were drawn from the pool of adult villagers on whom network data had been collected (that is, we did a full census of all adults in every village, mapped the whole network of each village as completely as possible, and then chose subjects from within this population, based on the algorithms below).

1. For randomly targeted villages, we drew a simple 5% random sample of these villagers.

2. For indegree-targeted villages, we targeted the 5% of villagers who had been named as a friend, sibling or spouse most often by others in their village. If tied values of indegree yielded more than 5% of the villagers, we randomly selected individuals who were tied, up to the 5% threshold. For instance, in a village with 100 eligible targets, the highest indegree individuals might have had indegree values of (i.e., may have received friendship/siblinghood/spouse nominations numbering): *15, 13, 12, 10, 10, 10, 10, 9*, etc. Since a 5% target group requires 5 individuals (of which three will be those with indegree values of 15, 13 and 12), how did we select among the four individuals with indegree values of 10? In this case, we would randomly draw two of the four individuals with indegree values of 10 to compose the target group (along with those individuals with values of 15, 13 and 12).

3. In the nominated-friends-targeted villages, we selected 5% of the population *at random*, using the same strategy used for the randomly chosen targets. However, instead of targeting these randomly selected individuals, we selected (again, *at random*) *one friend named by each of these individuals*, and used those *friends* as targets. Notably, because all participants in all villages were asked to nominate friends, targeted individuals in nomination-targeted villages were unaware they had been nominated, such that the effect of the targeting method reflects the structural positions of nominated friends alone, and not any additional psychosocial or "priming" effects of knowing that one had been nominated.

For villages receiving both interventions, we used a different targeting method for each intervention, and target groups were generated independently, such that 26 villagers (across all villages) were selected as targets for both interventions, by chance.

Choice of Outcome Measures

We chose ticket redemption as our *primary measure* of behaviour because it was the most accurately and comprehensively recorded measure of product uptake (i.e., we know the identity without exception of every

individual who redeemed a ticket, and the exact date on which he or she did so), and because it allowed us to trace with the greatest social and temporal resolution the rate and extent of product diffusion through the village networks, without relying on participants' recollection or self-report.

We supplemented this "hard" behavioural measure (ticket redemption) with self-reports of knowledge and practice as well, conducting an extensive follow-up survey in all villages in which we asked villagers (whether or not they had redeemed a ticket) about their use of the products, their attitudes concerning the products' utility and effectiveness, and a series of factual questions about their correct usage and benefits (from which the knowledge scores, our *secondary outcomes*, were derived, as reported in *Models of Knowledge Diffusion,* below).

Conditional on ticket redemption, there was no significant variation in self-reported product use or belief in the products' effectiveness by targeting method (random, indegree, or friend nomination). Self-reported continued product use among confirmed ticket redeemers was uniformly high (> 90%). And high multivitamin knowledge scores, as we report, were more common in friendship nomination-targeted villages.

These results suggest that ticket redemption (in which we *did* observe significant variation by targeting method, both at the population level, as well as in mixed-effects models accounting for individual-level and village-level covariates and for possible interference between interventions) is indeed a *valid omnibus measure of product adoption and continued use*, while the knowledge scores we report provide additional information about the differential spread of *health information* not predicted by product use alone.

Population-level Comparisons

As described in the main text, all pairwise differences in multivitamin redemption rates are statistically significant ($p < 0{\cdot}01$) after correction for multiple comparisons. The specific, pairwise, Holm-corrected p-values are as follows:

1. Nomination (951/1280 = 74·3% redeemed) vs. Random (940/1420 = 66·2% redeemed): $p < 0.0001$
2. Nomination (951/1280 = 74·3% redeemed) vs. Indegree (744/1220 = 61·0% redeemed): $p < 0.0001$
3. Random (940/1420 = 66·2% redeemed) vs. Indegree (744/1220 = 61·0% redeemed): $p = 0{\cdot}0062$

Ticket Redemption Models

For the multivitamin (MVI) intervention, the ticket-level model (Table S4) was estimated as follows. Of the 32 villages, 27 received MVI (24 of these received chlorine as well). 200 MVI targets were chosen across these 27 villages, resulting in 4,000 possible MVI tickets which could be redeemed (200 x 4 = 800 first-wave tickets, 800 x 4 = 3,200 second-wave tickets). Of these 4,000 tickets, 66% were redeemed by eligible participants. 52% of MVI ticket redeemers were also recorded as having redeemed a chlorine ticket. All dates are absolute (rather than zeroed at day of intervention) to control for day-of-week and other secular temporal effects.

Since MVI ticket redemption extended to 15 days after the introduction of the interventions in a given village, each unredeemed MVI ticket received 15 observations (ticket-days), for a total of 20,445 ticket-days for the 1,363 MVI tickets that remained unredeemed by the end of the study. For redeemed tickets, the number of ticket-days per ticket is equal to the number of days from initial targeting through redemption of that ticket.

We controlled for interference between the two interventions by including a ticket redeemer's redemption of a ticket for the opposite intervention as a time-varying covariate: an indicator for chlorine ticket redemption was set to 1 for any ticket-days on or after which the ultimate redeemer of the MVI ticket had redeemed a chlorine ticket. We include as ticket-level covariates the demographics (sex, age, household size, and marital status) of the targets associated with each ticket.

We used separate Cox proportional hazards models for first- and second-wave tickets to estimate the effect of targeting method on time-to-MVI-redemption, controlling for a subject's redemption of a ticket for the other intervention, as well as for basic demographics of the initial targets.

We did not include first-wave redeemer random effects because this would entail conditioning on post-treatment variables (i.e., the identity of the first-wave redeemers) and thus induce post-treatment bias.[3] Instead, we control for the characteristics of the *targets* associated with each ticket, since target selection (through random, indegree, or friendship nomination targeting) can be considered contemporaneous with or preceding "treatment" (i.e., introduction of the interventions to the villages), thus allowing us to control for potential confounders while preserving experimental causal inference.

We also controlled for basic village properties and blocking variables, and for block assignment. We accounted for clustering of observations within villages by estimating frailty models with random village intercepts (using the *coxme* package in R).

The models for chlorine ticket redemption (Table S5) were estimated in analogous fashion.

Models using generalised estimating equations (adjusting variance estimates for clusters of within-village-correlated observations based on a grouped jackknife, using the *geepack* package in R) produce qualitatively similar results.

Tables S4 and S5 show the beta coefficient estimates, hazard ratios (measured as the exponentiated coefficients), 95% confidence intervals for the hazard ratios, and *p* values.

Models of Knowledge Diffusion

Over the course of a single day for each village, we delivered to each targeted individual an intervention consisting of a health product (multivitamins or chlorine), instructions for use, and an educational component. Basic usage and safety information was repeated to ticket-redeemers when they received their products. However, we also gave the initially targeted individuals supplementary information about the interventions that was *not* generally known at baseline or circulated by other means, and asked them to relay it to those to whom they gave tickets, which allowed us to track the diffusion of knowledge as well as of product adoption by the study's completion.

For multivitamins, we taught targets to take 1 tablet per day. The supplementary usage information was to take the pill with food if the vitamin upsets an empty stomach. The educational information was that calcium strengthens bones, that iron prevents anemia, and that vitamin A aids vision.

For the chlorine intervention, recipients were taught to add 3 drops of bleach to 1 liter of tap water, stir, and leave sitting for 20 minutes before drinking. The supplementary usage information was to add 2 additional drops of chlorine if the water was cloudy. The educational information was that the correct use of chlorine kills the germs that cause diarrhea, which is especially harmful to young children.

Upon completion of the interventions, we returned to the 32 villages and administered a follow-up survey to all villagers, including targets, first- and second-wave ticket redeemers, and those who had neither been targeted nor had redeemed tickets. To assess the diffusion of knowledge about the interventions, we asked all non-targeted ticket recipients a series of questions about the use and benefits of the products.

For multivitamins, these questions included: how often to take the vitamins, what to do differently if the vitamin caused stomach upset, what vitamins were included in the multivitamin, and what health benefits were associated with the component vitamins. For chlorine, we asked ticket redeemers how many drops to use per liter of water, how long to let the water rest before consumption, what to do with cloudy water, what are the health consequences of contaminated water, and why purifying water is particularly important for children.

We formed composite 0-10 knowledge scores for each intervention, using the first component from a principal components analysis of the knowledge and usage questions posed during the follow-up survey. Since some of the information tested was introduced only to the original target individuals, the attainment of a high composite knowledge score (which we defined as a score in the top quartile) in an untargeted individual suggested that the usage and health information had diffused through the village network, starting from the target individuals. Because

the composite scores assumed a limited number of discrete values, the proportion of respondents in the top "quartile" is greater than 25%.

We estimated logistic regression models with random village intercepts, fit by maximum likelihood, for high knowledge score attainment among untargeted ticket recipients. We include the same individual-level and village-level covariates as in the ticket-redemption models, with the modification that, here, the individual-level covariates pertain to the ticket recipient herself, rather than to the original target individual. We control for possible interference between the two interventions by estimating the effects of the targeting methods of *both* products on the attainment of a high knowledge score for either product. Results for both interventions are presented in Table S6. Models using other measures of knowledge (including the raw composite score, rather than the PCA-derived score) produce qualitatively similar results (not shown).

Robustness of Results to Omitted Data

We omitted from village-level analysis (Figure 3 in main text) partial data from one village in which the ticket-redemption process was not accurately documented by the participating shopkeeper. This did not affect the main results: in the data as presented, multivitamin ticket redemption rates were 66·2% (940/1420) under random targeting, 61·0% (744/1220) under indegree targeting, and 74·3% (951/1280) under nomination targeting. With no data omitted, the respective rates are 65·9% (988/1500), 61·0% (744/1220) and 74·3% (951/1280).

For the chlorine intervention, aggregate ticket redemption rates as presented were statistically indistinguishable ($p > 0{\cdot}05$): 55·6% (690/1240) under random targeting, 55·9% (995/1780) under indegree, and 54·5% (512/940) under nomination targeting. With no data omitted, the respective rates are again statistically equivalent: 55·6% (690/1240), 55·9% (995/1780) and 53·4% (545/1020).

Tables

Table S1. Baseline survey results used for block assignment

Village number	Proportion of respondents drinking purified water	Proportion of respondents taking multivitamins daily	Village SES rating	Villagers with network data
1	0·38	0·11	5·9	120
2	0·15	0·02	3·3	43
3	0·10	0·06	4·3	39
4	0·10	0·05	7·4	108
5	0·33	0·15	7·4	254
6	0·49	0·10	1·4	141
7	0·41	0·12	6·0	180
8	0·49	0·10	6·4	90
9	0·07	0·03	2·4	130
10	0·77	0·22	6·1	240
11	0·60	0·10	4·4	368
12	0·21	0·10	1·8	112
13	0·21	0·17	5·6	369
14	0·72	0·08	1·3	88
15	0·33	0·11	2·0	72
16	0·19	0·09	4·1	112
17	0·18	0·10	3·4	38
18	0·62	0·21	6·0	25
19	0·13	0·04	2·8	88
20	0·70	0·12	8·0	35
21	0·29	0·07	1·3	60
22	0·98	0·13	3·4	174
23	0·51	0·11	5·0	301
24	0·67	0·22	8·4	348
25	0·17	0·14	6·6	101
26	0·34	0·14	3·9	206
27	0·56	0·10	2·4	151
28	0·14	0·07	4·0	79
29	0·95	0·04	1·0	64
30	0·33	0·10	6·4	390
31	0·15	0·10	4·7	72
32	0·78	0·07	4·1	64

Table S2. Randomised block experimental design

	Multivitamins				Chlorine			
Block	**Village 1**	**Village 2**	**Village 3**	**Village 4**	**Village 1**	**Village 2**	**Village 3**	**Village 4**
1	Random	Indegree	Nominated	Random	Indegree	Nominated	Random	None
2	Random	Indegree	Nominated	Indegree	Nominated	Random	Indegree	None
3	Random	Indegree	Nominated	Nominated	Indegree	Nominated	Random	None
4	Random	Indegree	Nominated	None	Nominated	Random	Indegree	Random
5	Random	Indegree	Nominated	None	Indegree	Nominated	Random	Indegree
6	Random	Indegree	Nominated	None	Nominated	Random	Indegree	Nominated
7	Random	Indegree	Nominated	None	Indegree	Nominated	Random	None
8	Random	Indegree	Nominated	None	Nominated	Random	Indegree	None

Table S3. Assignment of villages to block and targeting method

Village	Block	Multivit. Targeting	Chlorine targeting	Targets	Network size	Mean indegree of random targets (SD)	Mean indegree of nominated targets (SD)	Mean indegree of indegree targets (SD)
1	2	Rand	Nom	6	120	2·50 (2·26)	9·67 (4·59)	
2	5	Indeg	Nom	2	43		6·50 (2·12)	7·50 (3·54)
3	5	Nom	Rand	2	39	5·50 (4·95)	3·00 (0·00)	
4	6	Rand	Nom	5	108	4·20 (0·45)	3·80 (1·64)	
5	1	Indeg	Nom	13	254		5·54 (3·20)	8·58 (1·31)
6	2	Nom	Indeg	7	141		5·29 (3·25)	11·29 (2·29)
7	1	Nom	Rand	9	180	5·11 (3·33)	7·00 (4·87)	
8	7			5	90			
9	5		Indeg	7	130			10·86 (1·95)
10	4		Rand	12	240	4·83 (3·16)		
11	4	Indeg	Rand	18	368	4·67 (4·38)		13·17 (2·53)
12	6	Nom	Indeg	6	112		8·17 (4·71)	10·17 (2·79)
13	1	Rand	Indeg	18	369	5·94 (3·56)		11·50 (2·09)
14	7	Indeg	Nom	4	88		5·67 (1·53)	9·25 (3·20)
15	6	Indeg	Rand	4	72	4·00 (2·16)		8·00 (0·82)
16	6		Nom	6	112		9·67 (5·65)	
17	5	Rand	Indeg	2	38	1·50 (0·71)		7·50 (2·12)
18	7	Nom	Rand	1	25	2·00 (0·00)	3·00 (0·00)	
19	8	Rand	Nom	4	88	2·50 (2·38)	4·50 (1·73)	
20	3	Indeg	Nom	2	35		7·50 (6·36)	9·50 (3·54)
21	8	Indeg	Rand	3	60	5·33 (4·51)		9·67 (0·58)
22	4	Rand	Nom	9	174	3·11 (2·42)	6·89 (5·49)	
23	3	Nom		15	301		7·80 (5·52)	
24	4	Nom	Indeg	17	348		3·76 (2·36)	6·71 (1·76)
25	2	Indeg		5	101			9·80 (3·56)
26	2	Indeg	Rand	10	206	3·00 (2·36)		9·70 (2·11)
27	7	Rand	Indeg	8	151	4·12 (1·13)		12·75 (7·44)
28	8	Nom	Indeg	4	79		5·25 (1·50)	7·00 (0·82)
29	3	Nom	Rand	3	64	7·00 (1·00)	6·33 (3·06)	
30	3	Rand	Indeg	20	390	5·80 (3·76)		9·85 (2·16)
31	8			4	72			
32	1	Rand		3	64	2·33 (1·53)		

Table S4. Cox regression models with random village intercepts for multivitamin ticket redemption

	First wave tickets				*Second wave tickets*			
	Coef	**Hazard Ratio**	**95% CI (HR)**	***p***	**Coef**	**Hazard Ratio**	**95% CI (HR)**	***p***
Nomination targeting	**0·50**	**1·65**	**1·10, 2·47**	**0·02**	**0·14**	**1·15**	**0·74, 1·76**	**0·54**
Indegree targeting	**0·26**	**1·29**	**0·87, 1·93**	**0·20**	**0·07**	**1·07**	**0·70, 1·65**	**0·76**
Chlorine ticket redeemed	2·21	9·07	7·46, 11·03	0·00	2·50	12·23	10·87, 13·75	0·00
Target male	-0·34	0·71	0·59, 0·85	0·00	-0·12	0·89	0·80, 0·98	0·02
Target age	0·01	1·01	1·00, 1·02	0·01	0·00	1·00	1·00, 1·00	0·77
Target persons in house	0·02	1·02	0·98, 1·07	0·34	0·03	1·04	1·01, 1·06	0·01
Target married	-0·03	0·97	0·81, 1·16	0·75	-0·02	0·98	0·88, 1·08	0·63
Village mean indegree	0·11	1·12	0·80, 1·57	0·52	-0·09	0·91	0·63, 1·31	0·62
Village percent male	-0·03	0·97	0·93, 1·01	0·14	-0·04	0·96	0·92, 1·00	0·04
Village mean age	-0·03	0·97	0·88, 1·07	0·54	-0·01	0·99	0·89, 1·09	0·79
Village SES	-0·02	0·98	0·83, 1·17	0·86	-0·09	0·92	0·77, 1·10	0·34
Village population (100s)	-0·44	0·64	0·41, 1·01	0·06	-0·36	0·70	0·43, 1·14	0·15
Village baseline purification	-0·02	0·98	0·95, 1·02	0·35	-0·01	0·99	0·96, 1·02	0·50
Block 2	-0·53	0·59	0·26, 1·35	0·21	-0·19	0·83	0·32, 2·14	0·69
Block 3	0·71	2·04	0·96, 4·34	0·06	0·61	1·84	0·83, 4·04	0·13
Block 4	0·02	1·02	0·24, 4·44	0·98	0·10	1·10	0·23, 5·28	0·90
Block 5	-1·12	0·33	0·04, 2·60	0·29	-0·89	0·41	0·04, 3·90	0·44
Block 6	-1·83	0·16	0·03, 0·83	0·03	-1·39	0·25	0·04, 1·49	0·13
Block 7	-0·25	0·78	0·38, 1·58	0·49	0·08	1·08	0·52, 2·28	0·83
Block 8	-1·06	0·35	0·05, 2·57	0·30	-0·86	0·42	0·05, 3·63	0·43
Tickets redeemed	645				2000			
Random village intercept variance	0·08				0·15			
Deviance (null)	6758				28268			
Deviance (fitted)	6186				25996			

Table S5. Cox regression models with random village intercepts for chlorine ticket redemption

	First wave tickets				*Second wave tickets*			
	Coef	**Hazard Ratio**	**95% CI (HR)**	***p***	**Coef**	**Hazard Ratio**	**95% CI (HR)**	***p***
Nomination targeting	**0·06**	**1·06**	**0·60, 1·89**	**0·84**	**0·26**	**1·30**	**0·79, 2·14**	**0·31**
Indegree targeting	**-0·01**	**0·99**	**0·56, 1·75**	**0·97**	**-0·07**	**0·94**	**0·57, 1·53**	**0·79**
Multivitamin ticket redeemed	2·85	17·29	14·27, 20·95	0·00	2·99	19·98	17·70, 22·56	0·00
Target male	-0·01	0·99	0·82, 1·21	0·94	-0·08	0·92	0·82, 1·04	0·19
Target age	-0·01	0·99	0·99, 1·00	0·13	0·00	1·00	0·99, 1·00	0·21
Target persons in house	0·04	1·04	0·99, 1·10	0·12	-0·02	0·98	0·95, 1·01	0·20
Target married	-0·09	0·91	0·75, 1·10	0·32	-0·25	0·78	0·70, 0·87	0·00
Village mean indegree	0·55	1·74	1·13, 2·68	0·01	0·33	1·39	0·96, 2·03	0·08
Village percent male	0·02	1·02	0·96, 1·07	0·59	-0·03	0·97	0·93, 1·02	0·23
Village mean age	0·01	1·01	0·90, 1·14	0·82	0·01	1·01	0·92, 1·12	0·79
Village SES	0·21	1·24	0·94, 1·63	0·13	0·22	1·24	0·97, 1·59	0·08
Village population (100s)	0·25	1·28	0·49, 3·36	0·61	0·90	2·46	1·05, 5·75	0·04
Village baseline purification	0·02	1·02	0·95, 1·09	0·57	0·07	1·07	1·00, 1·14	0·04
Block 2	0·18	1·20	0·38, 3·82	0·76	0·65	1·91	0·67, 5·44	0·22
Block 3	0·05	1·05	0·26, 4·31	0·94	-1·32	0·27	0·08, 0·94	0·04
Block 4	-1·29	0·28	0·01, 6·13	0·42	-3·25	0·04	0·00, 0·62	0·02
Block 5	2·67	14·46	0·24, 875·70	0·20	4·32	74·83	1·90, 2948·67	0·02
Block 6	1·28	3·59	0·16, 82·45	0·42	2·92	18·52	1·13, 304·30	0·04
Block 7	0·68	1·98	0·73, 5·35	0·18	0·57	1·76	0·77, 4·04	0·18
Block 8	1·91	6·74	0·15, 307·47	0·33	4·01	55·20	1·85, 1646·42	0·02
Tickets redeemed	577				1625			
Random village intercept variance	0·11				0·11			
Deviance (null)	6292				23680			
Deviance (fitted)	5294				20348			

Table S6. Logistic regression models with random village intercepts for high knowledge score attainment

	Multivitamin high knowledge score				*Chlorine high knowledge score*			
	Coef	**Odds Ratio**	**95% CI (OR)**	***p***	**Coef**	**Odds Ratio**	**95% CI (OR)**	***p***
Nomination targeting (multivitamin)	**0·51**	**1·66**	**1·02, 2·70**	**0·04**	0·27	1·32	0·87, 1·98	0·19
Indegree targeting (multivitamin)	**0·23**	**1·25**	**0·84, 1·88**	**0·28**	0·10	1·10	0·76, 1·59	0·60
Nomination targeting (chlorine)	0·61	1·85	1·09, 3·14	0·02	**0·20**	**1·23**	**0·80, 1·88**	**0·34**
Indegree targeting (chlorine)	0·42	1·53	0·98, 2·38	0·06	**0·10**	**1·11**	**0·69, 1·79**	**0·67**
Male	0·02	1·02	0·84, 1·24	0·80	0·05	1·05	0·85, 1·30	0·63
Age	0·00	1·00	0·99, 1·01	0·73	0·00	1·00	0·99, 1·01	0·59
Persons in house	-0·02	0·99	0·93, 1·04	0·57	0·01	1·01	0·96, 1·07	0·69
Married	-0·06	0·94	0·77, 1·15	0·57	-0·14	0·87	0·71, 1·07	0·20
Village mean indegree	0·10	1·10	0·86, 1·41	0·44	0·16	1·17	0·91, 1·51	0·21
Village mean age	-0·11	0·90	0·82, 0·99	0·03	-0·04	0·96	0·90, 1·03	0·30
Village percent male	-0·02	0·98	0·94, 1·03	0·49	0·00	1·00	0·97, 1·04	0·84
Village SES	0·00	1·00	0·91, 1·11	0·92	0·05	1·06	0·95, 1·17	0·29
Village population (100s)	-0·06	0·95	0·76, 1·18	0·62	-0·17	0·85	0·66, 1·08	0·18
Village baseline purification	0·00	1·00	0·99, 1·01	0·72	-0·01	0·99	0·98, 1·00	0·05
Block 2	0·27	1·31	0·85, 2·02	0·21	-0·03	0·97	0·65, 1·44	0·87
Block 3	0·40	1·49	0·95, 2·33	0·08	0·25	1·28	0·75, 2·18	0·36
Block 4	0·38	1·46	0·70, 3·04	0·31	0·44	1·56	0·80, 3·01	0·19
Block 5	1·13	3·11	0·93, 10·35	0·06	0·01	1·01	0·45, 2·25	0·98
Block 6	0·09	1·09	0·54, 2·22	0·81	-0·50	0·61	0·34, 1·08	0·09
Block 7	0·02	1·02	0·58, 1·80	0·95	-0·15	0·86	0·45, 1·64	0·66
Block 8	0·44	1·56	0·72, 3·38	0·26	-0·45	0·64	0·31, 1·34	0·23
Intercept	2·56	12·97	0·15, 1152·94	0·26	0·30	1·35	0·05, 35·09	0·86
N (Individuals)	2110				1698			
N (Villages)	27				27			
Random village intercept variance	$1·80 \times 10^{-11}$				$1·75 \times 10^{-11}$			
Deviance (null)	2523				2222			
Deviance (fitted)	2481				2202			

References

1. Centola D. The Spread of Behavior in an Online Social Network Experiment. Science 2010;329:1194-7.
2. Centola D, Macy MW. Complex Contagions and the Weakness of Long Ties. American Journal of Sociology 2007;113:702-34.
3. King G, Zeng L. The Dangers of Extreme Counterfactuals. Political Analysis 2006;14:131-59.