RESEARCH ARTICLE

Behavioral dynamics of social propagation for weight loss: Evidence from a randomized controlled trial in a low-income Appalachian community

Daniel E. Zoughbie[1,2,3]*, Dillon Huddleston[2], Eric L. Ding[2,3]*

1 Institute of International Studies, University of California at Berkeley, Berkeley, California, United States of America, **2** Social Network Research Group, Microclinic International, San Francisco, California, United States of America, **3** Department of Public Health, New England Complex Systems Institute, Cambridge, Massachusetts, United States of America

* Eric.FeiglDing@necsi.edu (ELD); zoughbie@cal.berkeley.edu (DEZ)
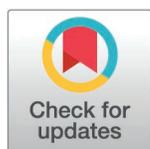
Abstract

Leveraging social propagation of health interventions for disease prevention is a public health holy grail. However, the dynamics of social causal induction have not been well studied. We present evidence from a randomized trial, demonstrating social-behaviorally "infectious" risk factor dynamics for obesity and diabetes. Specifically, we present the Social R_0 of weight loss propagation, which is itself adapted from the widely used infectious disease R_0. Via the trial design to infer causality, we calculated the Social R_0 across multiple time points for weight change propagation (*social-R_0* = 1.2, $p < 0.01$). We further show that health propagation of an intervention program can attain epidemic proportions; and that public health systems can intervene to modify the R_0 value thereby potentially managing, preventing, or reversing social infections at epidemic scale (yielding population-averaged *comparative R_0* ratios, $7.77 < R_c < 15.58$, $p < 0.01$, for weight loss propagation). To facilitate adoption of the methodologies, macros and code for use with various statistical software packages are included. The results indicate that the social induction of health interventions are not only possible, but that propagation can both be isolated causally via trial design and be quantified over time. This bodes enormous promise for developing and quantifying future self-sustaining public health interventions.

Author summary

Public health debate persists concerning the distinction between communicable diseases caused by microbes or viruses and "non-communicable diseases" that are not. We offer evidence that more precise and granular public health categories are needed; and that "non-communicable" and "non-infectious" can

Data availability statement: As the trial was conducted in small communities, the dataset contains potentially identifying or sensitive information. Even with anonymization, some risk of identification remains. Therefore, data

be tautological misnomers. The literature suggests an underlying socially "infectious" phenomenon and we hypothesize that social networks constitute a propagating force which can be quantified in socio-biological terms and leveraged for public health. But if weight loss, like biological pathogens, cascades through host populations, and can be mitigated through interventions targeting risky social behaviors, then can a social reproductive number for metabolic and cardiovascular disease be estimated and modulated? We present the first known evidence from a randomized trial in the USA, demonstrating the epidemiological dynamics of socially infectious risk factors. We find that health propagation can achieve epidemic proportions (including changes in weight); and that public health systems can intervene to potentially manage, prevent, or reverse social infections at epidemic scale. The results indicate that the social induction of health interventions are not only possible, but that propagation can be both isolated causally via trial design and quantified over time. This bodes enormous promise for developing and quantifying future self-sustaining public health interventions.

Introduction

Ever since Robert Koch's "germ theory" was proven in 1890 [1], a sharp distinction between diseases that are causally linked to microbes or viruses and those that are not has dominated the field of public health. The former are termed "infectious" or "communicable" and the latter "non-infectious" or "non-communicable." As the Cambridge Dictionary states, a non-infectious disease is "not able to be passed from one person, animal, or plant to another." An infectious disease is the reverse [2].

This binary logic sparked an unresolved debate within the WHO and in the pages that follow, we offer evidence in support of the view that fundamental public health categories need rethinking; that more precise and granular categories are needed; and that "non-infectious" and "non-communicable" can be tautological misnomers. If one assumes a disease is "non-infectious," one has no need to calculate a reproductive number (R_0). Post COVID-19, it is widely understood that the R_0 value, when acted upon strategically, can help public health officials pinpoint the location of an epidemic, estimate its severity, and deploy targeted preventative strategies. There is no end of lesson in comparing New Zealand's emergency response and associated COVID-19 R_0, which resulted in -57 excess deaths per million, and the United States' COVID-19 R_0, which resulted in 3,706 excess deaths per million [3].

Ever since Christakis and Fowler's 2007 observational study hinted there may be social correlations of obesity between limited friends [4], the scientific literature has been suggesting there may be an underlying socially "infectious" phenomenon. Due to inherent research design limitations in the Framingham Heart Survey [5], notably the inability to distinguish induction from homophily and confounding, causation could not be established. Russell Lyons questioned the study, even going so far as to suggest that the data may actually suggest no contagion [6]. Subsequent studies argued that ignoring homophily can inflate social network effects [7].

In the context of this debate, we put forth a fundamental Social Contagion Hypothesis: that social networks constitute a propagating force; that this force can be quantified in socio-biological terms; and that this force can be leveraged for public health. To prove this hypothesis, we first observed that weight loss trajectories clustered within social groups in Jordan [8]. We then designed two and three-arm randomized trials in rural Appalachia, Kentucky and in urban Amman, Jordan. As reported elsewhere, we demonstrated that social network effects enhanced educational effects, and using a novel ITT Social Induction Ratio, determined that social network effects explained the superior metabolic improvements of the main intervention. Our results were further bolstered by a mediator analysis: we found that behavioral improvements in diet, physical activity, medication, and monitoring did not substantially explain overall health improvements, suggesting the involvement of other causal pathways. We then identified a specific Leader-to-Follower Social induction pathway through which metabolic improvements cascaded and demonstrated social propagation within low-income Middle East populations [8–12]. As further validation of social network effects, the model was tested via randomized and quasi-experimental trials focused on HIV/AIDS in Kenya. These trials showed, in a vastly different infectious disease setting, that a social network intervention was causally linked to feelings of increased social support, and in a subpopulation analysis, demonstrated superior engagement in care [13].

Yet a major question still remains unanswered: if behavioral risk factors, like biological pathogens, cascade through host populations, and can be mitigated through interventions targeting risky social behaviors, then can a Social R_0 – an adaption of the infectious disease R_0 –for weight loss be estimated and modulated in a white rural low-income Appalachian setting?

In what follows, we present the first known evidence from a randomized trial, demonstrating the epidemiological dynamics of socially infectious diseases. We infer causality via a Social R_0 ratio; that an R_0 can be calculated for weight loss propagation; that health propagation can achieve epidemic proportions; and importantly, that public health systems can intervene to modify the Social R_0 value thereby potentially managing, preventing, or reversing social infections at epidemic scale.

Intervention

The Microclinic Social Network Program was initially designed as a game to test the possibilities and limitations of what we term "Managed Co-opetition." Since the overall structure of the program has been described elsewhere in great length [12,14–23] a brief overview is given below.

Potential program participants were screened and if they met the inclusion criteria were enrolled and randomized into two groups. In the first group, individuals formed "microclinic" clusters, small groups of family members or friends who supported one another in their journey towards improved health. Each of these clusters were nested within larger classes composed of other similar clusters.

During classes, microclinic clusters were given group activities and homework assignments to use their social influence to positively impact the health of those around them, especially those within their intimate circle of family and friends. They were taught a basic philosophy of the program, paraphrased as follows: *I influence and am influenced by those around me; this can be a good or bad thing; we can make this a good thing by making good health contagious!*

Participants were particularly instructed to apply this philosophy to four "M's" in their daily lives:

1. Meals (diet): "I influence and am influenced by the diet of those around me...."

2. Movement (physical activity): "I influence and am influenced by the physical activity of those around me...."

3. Monitoring (self-testing and testing at medical facilities): "I influence and am influenced by the testing of those around me...."

4. Medication (appropriate amounts, at the right time, in the right quantities). "I influence and am influenced by the medication consumption of those around me...."

Precisely because Microclinic clusters lived together or were good friends spending time supporting one another outside the study, they had the advantage of being able to organize themselves, to get their families to accept healthier meals, induce each other to exercise, keep tabs on one another's overall health and weight, and offer reminders to take their medication.

At the same time, there were clear elements of competition inherent in the program, occurring between Microclinic clusters in the same class, but also to a lesser extent, within the clusters. Importantly, a pedagogical aim made clear that there was a certain race to the top with the program, rather than a zero sum structure: one could help others improve their health, but one also needed the help of others in order to achieve their health goals. This would be equivalent, say, to a classroom full of math students being told that the grading curve would not only be abolished, but that they would also be given extra points if the class average was above a B + . In this scenario, each individual student would still compete for the prestige of an A+ mark, while at the same time, cooperating with others, keeping one another motivated, and reciprocally sharing information that may have been missed during lectures. We call this process "managed co-opetition."

In both competitive and cooperative cases, therefore, the payoff was tied to improved health, as was quantified via regular parallel testing among enrollees in both study arms.

Methodology

One novel extension of the Susceptible-Infected-Susceptible (SIS) compartmental model of mathematical epidemiology [24], the SISa model [25], augments the former with an additional, spontaneous, 'automatic' infection term:

$$\frac{dS}{dt} = -\beta SI + gI - aS = -\frac{dI}{dt}$$

Here, S (I) denotes the number of susceptible (infected) individuals at time t; a constant (large and well-mixed) population, $N = S + I$, is assumed; along with constant rates of transmission, recovery and spontaneous infection, respectively, β, g, and a. The authors estimate these rates, additionally assuming two kinds of infection ('content' and 'discontent') and allowing the possibility of superinfection directly between such states; epidemiologically modelling emotions in a social network determined via the Framingham study [5] which included administration of the CES-D exam [26] for classifying subjects' emotional states. We adapt their methodology to clinical, survey and social network data from a randomized controlled trial (RCT) conducted in Appalachian Bell County, Kentucky. A major difficulty is the classification of subjects' states: For changes in weight, we naturally identify 'lost' ('gained') with 'content' ('discontent'); but we have no CES-D or other diagnostic criteria to distinguish an intermediate, 'neutral' state. As such, much of the analysis and results center on sensitivity and robustness analyses of corresponding thresholds; absolute values of change below (above) which subjects are classified as 'neutral' ('lost' or 'gained').

Due to the supplementary spontaneous infection which distinguishes the SISa model, the basic reproduction ratio [27] does not exhibit the thresholding behavior common to classical compartmental models including SIS. Nonetheless, it may be estimated as follows:

$$R_0 \simeq \frac{\beta n}{g}$$

This is a first-order Taylor expansion of *1-exp(-βn/g)*, the cumulative distribution function of an exponential random variable corresponding to the interarrival time of a Poisson process with rate *βn/g*: It is the probability that, in a unit time interval (in the present case, one week), an 'infected' individual infects one of their (n) contacts. Here, n is the average network degree. However, there is an additional complication due to the model being constrained to a social network; the

estimated rates are specific to state transitions, and thus two R_0 values are computed (which measure the probabilities of such transitions), one each for (e.g., weight exceeding 5 lb) 'loss' and 'gain': For each $i \in \{l, g\}$,

$$\text{Social } R_{0,i} \simeq \frac{\beta_i n_i}{g_{T,i}}$$

Here, β_i is the slope estimated (a_i being the corresponding intercept) via an ordinary least squares (OLS) regression of neutral-to-state i transitions at time t on the number of contacts in state i at time t-1 (of which n_i is the average); and the 'total' rate,

$$g_{T,i} = g_i + s_i$$

Here, g_i is the intercept estimated via an OLS regression of state i-to-neutral transitions at time t on the number of neutral contacts at time t-1. Denoting by j the other infected state, and σ_{ij} the intercept estimated via an OLS regression of state i-to-state j transitions at time t on the number of contacts in state j at time t-1, the following are computed:

$$\alpha_j = a_j + \beta_j n_j$$

$$s_i = \left[\sigma_{ij} - g_i \left(1 - \frac{1 - e^{-\alpha_j}}{\alpha_j} \right) \right]_+$$

Here, $[.]_+$ denotes the positive part, $\max\{0,.\}$.

Our novel procedure involves three key R_0 causal intervention effect measures. First, having so computed $R_{0,l}$ and $R_{0,g}$, our 'integrated' *net loss* R_0 value is computed. Furthermore, in contrast to an observational study which was only able to establish the existence of dis/content R_0 values incidentally [25], the present RCT data was partitioned into two arms (treatment groups), treatment (A) and control (C), yielding values for each $x \in \{A, C\}$ and quantiles $q \in [0, 0.9]$:

$$R_{0,x;q} \equiv R_{0,x;l,q} - R_{0,x;g,q}$$

Here, for each $i \in \{l, g\}$, $R_{0,x;i,q}$ is as above, with contacts restricted to lie within common arms (x). This measures a sort of 'profit' realized within the corresponding arm; the probability of loss (improvement) less that of gain (dis-improvement).

More precisely, this weight loss net $R_{0,x;q}$ is interpreted as follows: Within arm $x \in \{A, C\}$, define the loss (gain) state via a previous-period weight decrease (increase) with magnitude exceeding quantile $q \in [0, 0.9]$ of the collection of all such one-period weight change magnitudes, and neutral otherwise; then $R_{0,x;q}$ is the *expected number of subsequent neutral-to-loss transitions induced by contacts in the loss state, less that of neutral-to-gain transitions induced by contacts in the gain state, both taken relative to a 'contactless' case where state transitions are not influenced by such contacts.* As such, either term (unlike a classical, epidemiological R_0) may be *negative*, in case of negative rather than positive influence, where loss (gain) contacts may be expected to in fact *decrease* the incidence of neutral-to-loss(-gain) transitions. E.g., heavy drug (say, heroine) use might be an intuitive example, where (hypothetically) having familial drug users may actually reduce one's likelihood of becoming a user. As a concrete example, in case $R_{0,x;l,q} = 2$ and $R_{0,x;g,q} = 0.5$, $R_{0,x;q} = 1.5$, say at a threshold quantile corresponding to 5 lb weight gain/loss, 1.5 more neutral-to-loss than -gain transitions are expected to occur, respectively due to loss-(gain-)state contacts. Also, since 1.5 > 1, over iterated time periods, such weight loss is expected to propagate as a contagion at an epidemic level.

Smoothed via moving average (with factor 0.75) and estimated [at the 95% level, propagating OLS coefficient estimates' standard errors via Monte Carlo simulations [28] and robustly estimating covariance via the Olive-Hawkins method

[29]], these values are obtained and plotted for thresholds (demarcating the neutral/infected state transitions) between zero and the 90th percentile of the histogram of absolute weight changes (for the entire dataset).

Second, to compare the two arms [which again is beyond an observational scope [25]], the following 'intent-to-treat causal *Social* R_0, relative efficacy ratio' is proposed:

$$R_{c;q} \equiv \frac{R_{0,A;q} - R_{0,C;q}}{|R_{0,C;q}|}$$

The numerator difference ensures that this ratio's sign corresponds to treatment efficacy; positive for treatment-induced weight loss. And the denominator absolute value scales this as a multiple of the control without regard to its sign. More precisely, the reason for using the absolute value of the denominator, is that the sign of the numerator then indicates whether the intervention improved upon the control (positive) or not (negative). These values are similarly smoothed and estimated for thresholds up to the 90th percentile of the histogram of absolute weight changes.

Third, a summary measure of the efficacies $R_{c;q}$, up to the qth percentile of the weight change histogram, is computed as a histogram [say, $p(\kappa)$]-weighted average of the preceding values:

$$\overline{R}_{c;q} = \int_0^q R_{c;\kappa} dp(\kappa)$$

Similarly, taking instead the arm-wise $R_{0,x;q}$ values as integrands, analogous summary measures are obtained:

$$\overline{R}_{0,x;q} = \int_0^q R_{0,x;\kappa} dp(\kappa)$$

Finally, note that the longitudinal, repeated-measures RCT data are spread across four one-week intervals; but given that there are too little data within any single of these, the state transitions and contact counts are pooled across all to yield significant results. Also, the OLS regressions are controlled for over a hundred covariates, including age, gender, preexisting health conditions such as pre/diabetes, hypertension and being overweight or obese, and several others obtained from survey data related to lifestyle habits including healthy eating, smoking and exercise.

To aid the reader's visualization of the mechanistic calculation of the Net Social R-naught, which represents the (Net of R(loss) – R(gain)), the steps of the previously described methods are diagramed in the four panels of Fig 1.

The multiple stepwise computations of Fig 1a, b are combined diagrammatically in Fig 2a; likewise for all of Fig 1 in Fig 2b for comparing the randomized trial's intervention arm versus the control arm, in order to obtain the causal effect efficacy ratio.

Results

The study included 494 participants randomized into a social network Intervention Group (n = 301) or the Control Group (n = 193). Recall the defined weight loss net $R_{0,x;q}$: Within arm $x \in \{A, C\}$, define the loss (gain) state via a previous-period weight decrease (increase) with magnitude exceeding quantile $q \in [0, 0.9]$ of the collection of all such one-period weight change magnitudes, and neutral otherwise; then $R_{0,x;q}$ is the *expected number of subsequent neutral-to-loss transitions induced by contacts in the loss state, less that of neutral-to-gain transitions induced by contacts in the gain state, both taken relative to a 'contactless' case where state transitions are not influenced by such contacts.* As a concrete example, in case $R_{0,x;l,q} = 2$ and $R_{0,x;g,q}$=0.5, $R_{0,x;q} = 1.5$, say at a threshold quantile corresponding to 5 lb weight gain/loss, 1.5 more neutral-to-loss than -gain transitions are expected to occur, respectively due to loss-(gain-)state contacts. Also, since 1.5 > 1, over iterated time periods, such weight loss is expected to propagate as a contagion at an epidemic level.
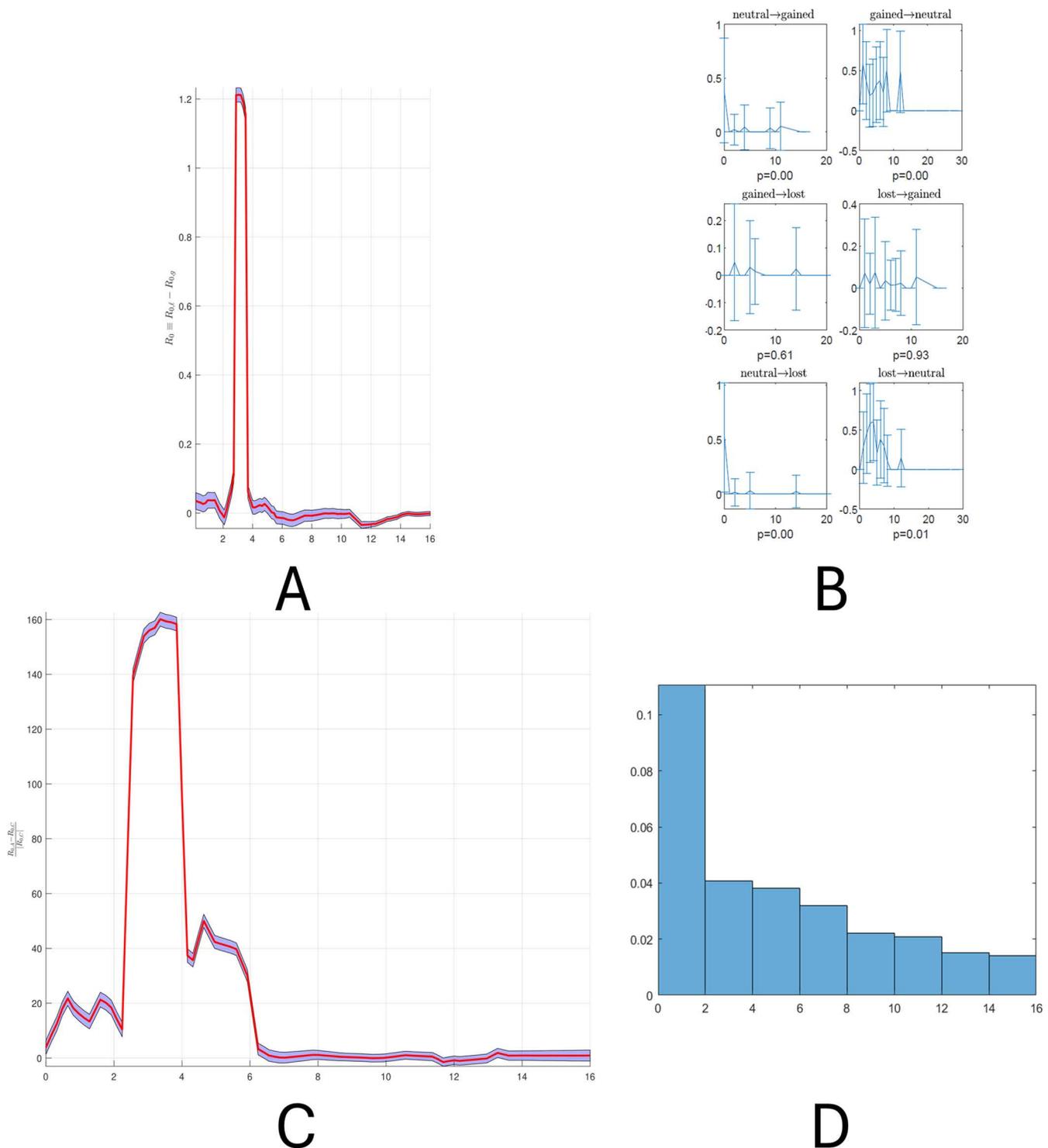
Fig 1. A. Specify a threshold (*x*-axis) for weight change above (below) which absolute value positive/negative weight changes are classified as R-naught(gain) and R-naught(loss) from a neutral level. B. Perform ordinary least squares regression on each of six non-self transitions between these three classes, for each subject eligible to so transition against their number of contacts likewise transitioning in the previous period. Coefficients from the four transitions involving loss (gain) are used to compute $R_{0,l}$ ($R_{0,g}$), which difference yields R_0 (previous panel *y*-axis). C. Performing the

preceding steps for the treatment arm and the control arms separately yields $R_{0,A}$ ($R_{0,C}$), respectively, which difference divided by the absolute value of the latter yields R_c. D. Numerically integrating R_c up to the qth percentile of the histogram of all subjects' weight changes (shown), weighted by the latter, yields the so-weighted average, $\overline{R}_{c,q}$ or the overall Net Social R-naught(weighted).

Table 1 reports trial participants' baseline characteristics.

Table 2 reports the histogram-weighted averages of the proposed social ratios (R_0) for weight loss.

Similar to taking net loss less gain weight changes, net intervention against control differencing establishes treatment effect. Provided are point and interval estimates for these weight loss net $R_{0,x;q}$ values, averaged over quantiles up to $\overline{q} = 0.9$ and weighted by the (within-arm) histogram of absolute weight changes: Intervention [control] $\overline{R}_{0,90}$ weight loss is estimated to be 0.0916 (95% CI [0.0686,0.115]; $p < 0.001$) [0.138 (95% CI [-0.0044,0.280]; $p = 0.058$)].

As to why the causal effect ratio reported in the third column of Table 2 doesn't equal or even approximate the naïve crude ratio of the values in the intervention versus control arms:

$$\overline{R}_{c;q} = \int_0^q R_{c;\kappa}\,dp(\kappa) = \int_0^q \frac{R_{0,A;\kappa} - R_{0,C;\kappa}}{|R_{0,C;\kappa}|}\,dp(\kappa) \neq \frac{\int_0^q R_{0,A;\kappa}\,dp(\kappa)}{\int_0^q R_{0,C;\kappa}\,dp(\kappa)} = \frac{\overline{R}_{0,A;q}}{\overline{R}_{0,C;q}}$$

The formula is more than just a crude ratio – as mathematical rules dictate that "the integral of a ratio is not the ratio of the integrals:" Both because $R_{c;\kappa}$ divides by the *absolute* value $|R_{0,C;\kappa}|$, and the numerator difference without that being the case would result in subtracting q from the left-hand equalities. We believe our way is superior to a straightforward crude ratio as the right-hand pair of equalities: The latter compares averages over the appropriate thresholds which 'blurs' their effect, whereas the former directly compares arms for each threshold, before averaging which avoids 'apples-to-oranges' comparisons. Intuitively, this avoids missing important facts, e.g., that peaks in the S1 Fig occur for different thresholds; for which the intervention-vs.-control comparison is strong, and remains so even when averaged over all thresholds; whereas averaging before taking ratios smoothes the peaks and ends up artificially comparing distributional averages as though they correspond to equal thresholds, hence the comparatively unimpressive ratios that result from dividing the first two columns of Table 2.

Table 3 reports the histogram-weighted averages of the proposed (*comparative*) ratios (R_c) for weight loss.

Provided are point and interval estimates for these weight loss net $R_{0;q}$ treatment effects (divided by the corresponding absolute control loss net values), averaged over quantiles up to $\overline{q} \in \{0.75, 0.8, 0.9\}$ and weighted by the histogram of absolute weight changes; with results seen to be consistent, and in particular for $\overline{q} = 0.9$ (replicated as the rightmost column of Table 2): $\overline{R}_{c,90}$ weight loss is estimated to be 13.7 (95% CI [10.6,16.7]; $p < 0.001$).

S1 Fig presents for weight loss, plots of R_0 by arm and R_c up to the 90th percentile of changes across the dataset; the histogram-weighted averages of the latter being tabulated previously.

Discussion

These results demonstrate the positive propensity for the propagation of weight loss in classrooms over time; cascading behavior that public health authorities can leverage. In a rural low-income setting, we tested novel Social R_0 metrics for the effect of behavioral change interventions on transmitting by weight changes over time. This was done loss net of gain for thresholds up to 90% of the population changes observed for weight loss, which has never been so parameterized before nor applied in the ITT setting of an RCT to estimate R_0 efficacy. Further, this integrated measure allows results to be translated across a range of weight loss thresholds. These are possible only because of our unique RCT design and the formulas presented. We believe these results suggest that social networks, and not simply health education, yield propagative clinical improvements. Participants exposed to the intervention experienced greater clinical improvements
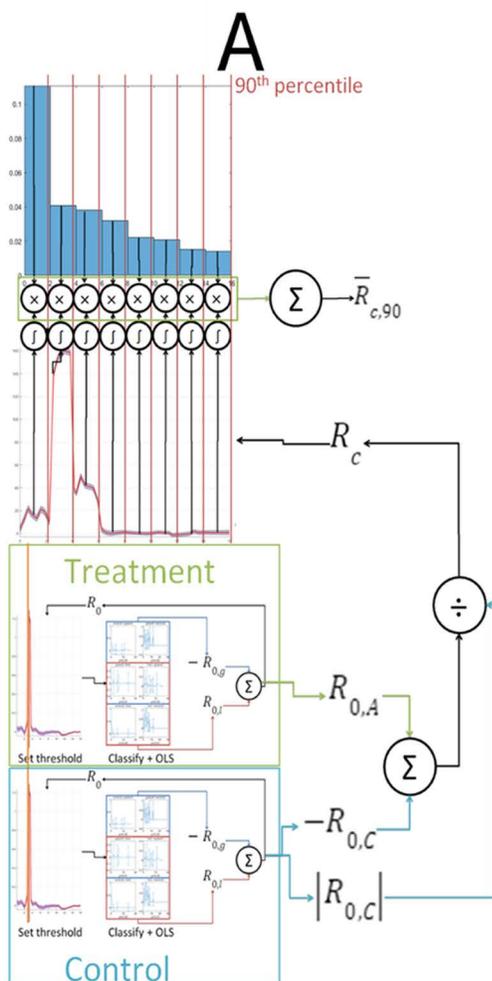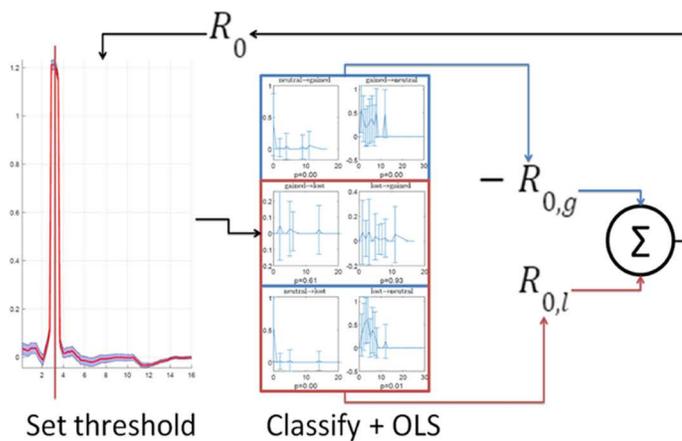
Fig 2. A. Net *R0(Loss-Gain)* calculation: For a given threshold, subjects are classified as gained, lost or neutral; OLS regression is done for each of the corresponding pairwise transitions; and those not involving gained (lost) are combined to yield $R_{0,l}$ ($R_{0,g}$); the difference of which yields Net *R0(Loss-Gain)* at each weight change threshold. B. This procedure is separately applied to treatment and control subjects, respectively

yielding $R_{0,A}$ and $R_{0,C}$; the difference of which, divided by the absolute value of the latter, yields R_c; which is then numerically integrated with respect to the histogram of (all subjects') weight changes to yield $\overline{R}_{c,q}$, e.g., in case $q = 90$, up to the 90th percentile.

Table 1. Baseline characteristics of Kentucky trial participants.

Characteristic, mean$\pm$SD, or %	Intervention Group (n = 301)	Control Group (n = 193)
Age, years	52.7$\pm$13.7	56.6$\pm$13.3
Women, (%)	88.4	84.1
Weight, kg	94.7$\pm$20.0	95.7$\pm$21.6
Height, m	1.63$\pm$	1.64$\pm$
BMI, kg/m^2	36.2$\pm$7.4	36.0$\pm$8.0
Obesity, BMI>=30, %	78.2	76.5
Waist circumference, inches	44.1$\pm$5.5	44.5$\pm$5.7
HbA1c %	6.1	6.2
HDL cholesterol, mg/dl	42.3$\pm$13.2	38.9$\pm$12.2
Mean arterial blood pressure, mmHg	95.6$\pm$10.4	96.0$\pm$10.2

Table 2. Causal Effect Ratio of Social R_0 Reproductive Number (0-90th percentile) in Classrooms over all intervention and control sessions of the randomized trial.

	Intervention Arm Social $\overline{R}_{0,A,90}$	Control Arm Social $\overline{R}_{0,C,90}$	Causal effect: Ratio of Social $\overline{R}_{c,90}$
Body Weight	0.0916 [0.0686,0.115]	0.138 [-0.0044,0.280]	13.7 [10.6,16.7]

Table 3. Body Weight histogram threshold-weighted average R_c up to 75th, 80th, and 90th percentiles.

(Net of R(loss) – R(gain))	Histogram-weighted average Rc (up to the listed percentile), $\overline{R}_{c,q}$		
q	0 to 75th percentile	0 to 80th percentile	0 to 90th percentile
Body Weight	15.6 [12.9,18.3]	7.77 [4.86,10.7]	13.7 [10.6,16.7]

relative to controls and the cascading effect of the Social R_0 is orders of magnitude larger. Because the study was randomized and results are consistent with other evidence, we believe we can rule out confounding factors or reverse causation.

Note that social reproductive numbers have been proposed and used widely in the case of biologically infectious communicable diseases, namely, for COVID-19 [30]. Frontier methods including convergent cross mapping and other instances of empirical dynamic modeling and manifold learning can account for variable non-linear interactions (vs. simple correlation) even without considering time delays [31]. These models are also able to capture important spatial dependencies, as well as variable collective interactions (i.e., multi-correlation); however, the size of the RCT dataset presently considered could be one limiting factor in their implementation. Similarly hampered are global sensitivity and uncertainty analysis (GSUA), via both simple variance-based and entropic approaches, which ideally could assess how much variability (and its randomness) is contained in inputs for the variability of outputs (or even distributionally via co-predictability versus causality) [32,33]. Altogether to provide an aspirational summary, in an ideal scenario with unlimited RCT data; how indicators/predicted variables evolve spatiotemporally, conditional on desired (health) outcomes, is critical to understand

site-/time-specific and universal shifts but more importantly the shape of stress-response patterns (e.g., how outcomes can be controlled by using the most important factors derived from GSUA). Indicator distributions (of predicted patterns) can be analyzed as a function of the predictors considering their joint distributions, or moments and considering indicator variability along predictors' gradients. The stability of such patterns over predictors' gradients and their critical transitions, is important to quantify because it may determine potential stable states.

In presenting these results for an intervention aiming to modulate the Social R_0 of so-called "non-infectious" or "non-communicable diseases," we show that the latter in fact have "socially infectious" or "socially communicable" properties that shape biology. While prior randomized and non-randomized trials have demonstrated varying levels of health improvements as a result of social-network structured interventions, they were not able to disentangle precise causal mechanisms [34–37]. By contrast, our studies, which have been replicated in two distinct socio-economic and geographical contexts, are the first to specifically isolate social networks as an independent causal force enacting, and being enacted upon, by individual biology [12]. In both quantifying and modulating social infection vis-a-vis a novel Social R_0 ratio, we have identified the essential socio-biological nexus of weight loss transmission. This bears great significance for the fields of epidemiology, implementation science, integrative biology, and especially public health in the age of artificial intelligence [38–40]. Notably, we have clarified the *interdependence* of individual and collective biological organisms in a "non-communicable" disease context lacking a physical pathogen.

Supporting information

S1 Fig. A. Weight change; R_0 values by arm: Most notable here is an *epidemic* value of R_0, loss net gain, in arm A: Separately from the 'automatic' infection (term *a*) which distinguishes the SISa model from its classical compartmental counterpart, SIS, it is significantly (at the 95% level, and likewise for all following results) observed in the trial (treatment) arm A that for every subject having lost between 2.8-3.6 lb or more, on average in the following week, more than one (specifically, about 1.2) of their previously neutral social contacts also lose between 2.8-3.6 lb or more, *relative to the corresponding number of previously neutral social contacts that gain between 2.8-3.6 lb or more, for every subject having gained between 2.8-3.6 lb or more in the preceding week.* E.g., this situation could involve each 'loser' inducing 1.2 loser contacts and each 'winner' none, each loser and winner inducing 0.6 loser and winner contacts, or each winner inducing 1.2 winner contacts and each loser none: However, *net of gain*, each loser induces on average 1.2 loser contacts. No other comparable R_0 values are observed in arm A, whereas in (control) arm C, significant but smaller (and in particular, *non*-epidemic) net values are observed for a broader range of thresholds. To emphasize, relative to control (arm C), treatment (arm A) has significantly and for a range of change thresholds, induced positive, *epidemic* weight loss net of gain. B. R_c values: Notable here is an anticipated jump for the same thresholds which yield epidemic social-network-induced weight loss net gain; namely, 2.8-3.6 lb: The conclusion is that relative to control arm C, the differential improvement in treatment arm A is (significantly) two orders of magnitude greater than the absolute R_0 of weight loss net gain. In other words, not only does treatment induce epidemic-level improvements in weight loss from the control, but as a multiple of the latter's magnitude, the degree of improvement is nearly 250 times. Of course, there is a similarly large *negative* deviation from the control for much higher thresholds, which are seen in the corresponding R_0 plots to result from a much smaller (by about two orders of magnitude and obviously non-epidemic, but negative) R_0 in treatment arm A; but this is of little importance: Namely, because for such thresholds which exceed even the mean weight change in the entire dataset, the analysis becomes driven by outliers exhibiting extreme weekly weight loss or gain (all others being classified as neutral); and as such the histogram-weighted average R_c value, 13.66, is significantly positive as it heavily discounts these outlying aberrations.
(TIF)

S1 File. Pages 1–2: Macro code for computing the Social R0 in Matlab. Pages 3–4: Macro code for computing the Social R0 in R. Pages 5–6: Macro code for computing the Social R0 in STATA.
(DOCX)

Author contributions

Conceptualization: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Data curation: Daniel Zoughbie, Eric L. Ding.

Formal analysis: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Funding acquisition: Daniel Zoughbie, Eric L. Ding.

Investigation: Daniel Zoughbie, Eric L. Ding.

Methodology: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Project administration: Daniel Zoughbie, Eric L. Ding.

Resources: Daniel Zoughbie, Eric L. Ding.

Software: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Supervision: Daniel Zoughbie, Eric L. Ding.

Validation: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Visualization: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Writing – original draft: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

Writing – review & editing: Daniel Zoughbie, Dillon Huddleston, Eric L. Ding.

References

1. Blevins SM, Bronze MS. Robert Koch and the "golden age" of bacteriology. Int J Infect Dis. 2010;14(9):e744–51. https://doi.org/10.1016/j.ijid.2009.12.003 PMID: 20413340

2. Non-infectious. Cambridge English Dictionary. [Accessed 2023 October 14]. https://dictionary.cambridge.org/us/dictionary/english/non-infectious

3. Wang H. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. The Lancet. 2022;399(10334):1513–36.

4. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357(4):370–9. https://doi.org/10.1056/NEJMsa066082 PMID: 17652652

5. Dawber TR. The Framingham study: the epidemiology of atherosclerotic disease. Harvard University Press; 1980.

6. Lyons R. The spread of evidence-poor medicine via flawed social-network analysis. Stat Pol Policy. 2011;2(1). https://doi.org/10.2202/2151-7509.1024

7. Davin JP, Gupta S, Piskorski MJ. Separating homophily and peer influence with latent space. 2013. https://ssrn.com/abstract=2373273

8. Zoughbie DE, Rushakoff JA, Watson KT, Bui N, Ireifij A, Farraj RS, et al. A social-network behavioral health program on sustained long-term body weight and glycemic outcomes: 2-year follow-up of a 4-month Microclinic Health Program in Jordan. Prev Med Rep. 2018;13:160–5. https://doi.org/10.1016/j.pmedr.2018.12.002 PMID: 30596007

9. Zoughbie DE, Watson KT, Bui N. Tackling diabetes as a 'communicable' disease. In: From theory to practice: an in-depth look at three translational research projects supported by IDF Bridges first round of funding. Montreal; 2009.

10. Ding EL, Feigl AB, Watson KT, Ng TLJ, Makerechi L, Bui N, et al. Social network enhanced behavioral interventions for diabetes and obesity: a 3 arm randomized trial with 2 years follow-up in Jordan. PLOS Glob Public Health. 2024;4(3):e0001514. https://doi.org/10.1371/journal.pgph.0001514 PMID: 38507441

11. Salmen CR, Hickey MD, Fiorella KJ, Omollo D, Ouma G, Zoughbie D, et al. "Wan Kanyakla" (We are together): community transformations in Kenya following a social network intervention for HIV care. Soc Sci Med. 2015;147:332–40. https://doi.org/10.1016/j.socscimed.2015.11.021 PMID: 26638145

12. Zoughbie DE, Huddleston D, Ding EL. Social induction dynamics of the causal social R0 on clinical weight loss: randomized trial evidence of social propagation from Amman, Jordan. PLOS Complex Syst. 2025;2(2):e0000031. https://doi.org/10.1371/journal.pcsy.0000031

13. Hickey MD, Ouma GB, Mattah B, Pederson B, DesLauriers NR, Mohamed P, et al. The Kanyakla study: randomized controlled trial of a microclinic social network intervention for promoting engagement and retention in HIV care in rural western Kenya. PLoS One. 2021;16(9):e0255945. https://doi.org/10.1371/journal.pone.0255945 PMID: 34516557

14. Ding EL, Watson KT, Makarechi L, James Ng TL, Huddleston D, Bui N, et al. Social induction via a social behavioral intervention on changes in metabolic risk factors: a randomized controlled trial in rural Appalachia, United States. Mayo Clin Proc. 2024;99(7):1058–77. https://doi.org/10.1016/j.mayocp.2023.11.023 PMID: 38960495

15. Hickey MD, Salmen CR, Omollo D, Mattah B, Fiorella KJ, Geng EH, et al. Implementation and operational research: pulling the network together: quasiexperimental trial of a patient-defined support network intervention for promoting engagement in HIV care and medication adherence on Mfangano Island, Kenya. J Acquir Immune Defic Syndr. 2015;69(4):e127-34. https://doi.org/10.1097/QAI.0000000000000664 PMID: 25984711

16. Rushakoff JA, Zoughbie DE, Bui N, DeVito K, Makarechi L, Kubo H. Evaluation of Healthy2Go: a country store transformation project to improve the food environment and consumer choices in Appalachian Kentucky. Prev Med Rep. 2017;7:187–92. https://doi.org/10.1016/j.pmedr.2017.06.009 PMID: 28706778

17. Zoughbie DE. Community-based diabetes programme: the micro-clinic project. East Mediterr Health J. 2009;15(4):1021–6. https://doi.org/10.26719/2009.15.4.1021 PMID: 20187555

18. Shahin YM, Kishk NA, Turki Y, Saleh S, Ding EL, Zoughbie DE, et al. Evaluation of the microclinic social network model for palestine refugees with diabetes at UNRWA health centers. JDM. 2018;08(04):99–113. https://doi.org/10.4236/jdm.2018.84010

19. Prescott M, Zoughbie D, Watson K, Bui N, Farraj R, Elkarra N. The microclinic health program: a social network-based intervention for weight loss and diabetes risk management. Am J Epidemiol. 2013;177:S96.

20. Zoughbie DE, Ng TLJ, Thompson JY, Watson KT, Farraj R, Ding EL. Ramadan fasting and weight change trajectories: time-varying association of weight during and after Ramadan in low-income and refugee populations. PLOS Glob Public Health. 2022;2(10):e0000371. https://doi.org/10.1371/journal.pgph.0000371 PMID: 36962504

21. Zoughbie DE, Huddleston D, Ding EL. Game theory and weight loss: harmonic evidence from randomized controlled trials in Appalachia, KY and Jordan. PLOS Glob Public Health. 2025;5(2):e0004100. https://doi.org/10.1371/journal.pgph.0004100 PMID: 39919098

22. Zoughbie DE, Huddleston D, Watson K, Ding EL. HIV Social-network intervention more effective in older populations in Kenya. BMC Public Health. 2024;24(1):3098. https://doi.org/10.1186/s12889-024-20315-0 PMID: 39516844

23. Zoughbie DE, Ding EL, Ng TLJ. "Follow-the-Leader" domino weight loss effects in low-income Middle Eastern refugee communities: disentangling macro average peer versus micro leader-driven effects in a randomized trial. PLOS Complex Syst. 2025;2(7):e0000052. https://doi.org/10.1371/journal.pcsy.0000052

24. Hethcote HW. Three basic epidemiological models. Applied mathematical ecology. Berlin, Heidelberg: Springer Berlin Heidelberg; 1989: 119–44.

25. Hill AL, Rand DG, Nowak MA, Christakis NA. Emotions as infectious diseases in a large social network: the SISa model. Proc Biol Sci. 2010;277(1701):3827–35. https://doi.org/10.1098/rspb.2010.1217 PMID: 20610424

26. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measur. 1977;1(3):385–401.

27. Van den Driessche P, Watmough J. Further notes on the basic reproduction number. Math Epidemiol. 2008:159–78.

28. Kroese DP, Taimre T, Botev ZI. Handbook of monte carlo methods. John Wiley & Sons; 2013.

29. Olive DJ. A resistant estimator of multivariate location and dispersion. Comput Stat Data Ana. 2004;46(1):93–102. https://doi.org/10.1016/s0167-9473(03)00119-1

30. Chan LYH, Yuan B, Convertino M. COVID-19 non-pharmaceutical intervention portfolio effectiveness and risk communication predominance. Sci Rep. 2021;11(1):10605. https://doi.org/10.1038/s41598-021-88309-1 PMID: 34012040

31. Sugihara G, May R, Ye H, Hsieh C, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. Science. 2012;338(6106):496–500. https://doi.org/10.1126/science.1227079 PMID: 22997134

32. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, et al. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ Modell Soft. 2016;79:214–32. https://doi.org/10.1016/j.envsoft.2016.02.008

33. Servadio JL, Convertino M. Optimal information networks: application for data-driven integrated health in populations. Sci Adv. 2018;4(2):e1701088. https://doi.org/10.1126/sciadv.1701088 PMID: 29423440

34. Hunter RF, de la Haye K, Murray JM, Badham J, Valente TW, Clarke M, et al. Social network interventions for health behaviours and outcomes: a systematic review and meta-analysis. PLoS Med. 2019;16(9):e1002890. https://doi.org/10.1371/journal.pmed.1002890 PMID: 31479454

35. Zhang S, de la Haye K, Ji M, An R. Applications of social network analysis to obesity: a systematic review. Obes Rev. 2018;19(7):976–88. https://doi.org/10.1111/obr.12684 PMID: 29676508

36. Loh YL, Yaw QP, Lau Y. Social media-based interventions for adults with obesity and overweight: a meta-analysis and meta-regression. Int J Obes (Lond). 2023;47(7):606–21. https://doi.org/10.1038/s41366-023-01304-6 PMID: 37012428

37. Fu F, Christakis NA, Fowler JH. Dueling biological and social contagions. Sci Rep. 2017;7:43634. https://doi.org/10.1038/srep43634 PMID: 28252663

38. Jozani M, Ebadi N, Khan R, Zoughbie D. ChatGPC: a novel interactive AI health coach for obesity and diabetes prevention. 2025.

39. Bjerregaard AA, Zoughbie DE, Hansen JV, Granström C, Strøm M, Halldórsson ÞI, et al. An SMS chatbot digital educational program to increase healthy eating behaviors in adolescence: a multifactorial randomized controlled trial among 7,890 participants in the Danish National Birth Cohort. PLoS Med. 2024;21(6):e1004383. https://doi.org/10.1371/journal.pmed.1004383 PMID: 38875292

40. Ackland M, Choi BCK, Puska P. Rethinking the terms non-communicable disease and chronic disease. J Epidemiol Community Health. 2003;57(11):838–9. https://doi.org/10.1136/jech.57.11.838 PMID: 14600102