

# BMJ Open Diffusion of an evidence-based smoking cessation intervention through Facebook: a randomised controlled trial study protocol

Nathan K Cobb,<sup>1,2</sup> Megan A Jacobs,<sup>3</sup> Jessie Saul,<sup>4</sup> E Paul Wileyto,<sup>5</sup> Amanda L Graham<sup>3,6</sup>

**To cite:** Cobb NK, Jacobs MA, Saul J, *et al*. Diffusion of an evidence-based smoking cessation intervention through Facebook: a randomised controlled trial study protocol. *BMJ Open* 2014;**4**: e004089. doi:10.1136/bmjopen-2013-004089

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-004089>).

Received 20 September 2013  
Revised 5 December 2013  
Accepted 12 December 2013



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Dr Amanda L Graham;  
[agraham@legacyforhealth.org](mailto:agraham@legacyforhealth.org)

## ABSTRACT

**Introduction:** Online social networks represent a potential mechanism for the dissemination of health interventions including smoking cessation; however, which elements of an intervention determine diffusion between participants is unclear. Diffusion is frequently measured using R, the reproductive rate, which is determined by the duration of use (t), the ‘contagiousness’ of an intervention ( $\beta$ ) and a participant’s total contacts (z). We have developed a Facebook ‘app’ that allows us to enable or disable various components designed to impact the duration of use (expanded content, proactive contact), contagiousness (active and passive sharing) and number of contacts (use by non-smoker supporters). We hypothesised that these elements would be synergistic in their impact on R, while including non-smokers would induce a ‘carrier’ state allowing the app to bridge clusters of smokers.

**Methods and analysis:** This study is a fractional factorial, randomised control trial of the diffusion of a Facebook application for smoking cessation. Participants recruited through online advertising are randomised to 1 of 12 cells and serve as ‘seed’ users. All user interactions are tracked, including social interactions with friends. Individuals installing the application that can be traced back to a seed participant are deemed ‘descendants’ and form the outcome of interest. Analysis will be conducted using Poisson regression, with event count as the outcome and the number of seeds in the cell as the exposure.

**Results:** The results will be reported as a baseline R0 for the reference group, and incidence rate ratio for the remainder of predictors.

**Ethics and Dissemination:** This study uses an abbreviated consent process designed to minimise barriers to adoption and was deemed to be minimal risk by the Institutional Review Board (IRB). Results will be disseminated through traditional academic literature as well as social media. If feasible, anonymised data and underlying source code are intended to be made available under an open source license.

**ClinicalTrials.gov registration number:** NCT01746472.

## INTRODUCTION

Smoking remains the leading cause of 443 000 preventable deaths and US\$200 billion in excess cost in the USA each year,<sup>1</sup> making a large-scale reduction in smoking prevalence a public health imperative. Yet, evidence-based interventions recommended by the Clinical Practice Guideline for Tobacco Dependence Treatment (‘2008 Guideline’)<sup>2</sup> do not reach the vast majority of the 44 million current smokers in the USA.<sup>3–5</sup> A major paradigm shift in how cessation interventions are developed is needed, targeting a large-scale dissemination and diffusion.<sup>2</sup>

In theory, the broad reach and effectiveness of evidence-based Internet cessation programmes should yield enormous impact (reach  $\times$  efficacy<sup>6</sup>) in reducing the population prevalence of smoking. The majority (85%) of US adults are Internet users, including populations at disproportionate risk for smoking: 85% of African Americans and 76% of those with incomes less than US\$30 000/year use the Internet.<sup>7</sup> Between 6% and 9% of all Internet users (>10 million adults) search for quitting smoking annually.<sup>8, 9</sup> Studies and multiple meta-analyses<sup>10–13</sup> show that Internet interventions are effective with a relative risk of abstinence of 1.44<sup>10</sup> and quit rates of 7–26%.<sup>14–17</sup> Despite this promise, however, only one-third of smokers searching the Internet actually reach the limited number of websites that provide cessation treatment consistent with the 2008 Guideline.<sup>18–20</sup> Most existing Internet cessation interventions that involve social support—a key element of tobacco dependence treatment—introduce participants into ‘artificial’ networks in which individuals have no initial connections and often create none. In such networks, participation is limited by affiliation with a particular behaviour (eg, quitting smoking). As a result, potential network effects on individual behaviour and the potential for

dissemination are sharply limited by high levels of attrition,<sup>21</sup> the fact that most registrants never form a single connection,<sup>22</sup> and those that are formed may be weak and transitory.

Online social networks may represent a more powerful dissemination channel for evidence-based tobacco dependence treatments. In contrast to the 'build it and they will come' model inherent in smoking-specific online cessation interventions, more general online social networks can be used to deliver proven cessation intervention elements to smokers 'where they are'. Two-thirds (67%) of the US adult Internet users use at least one social networking site such as Facebook or Twitter; importantly, nearly 80% of adults aged 18–49 do so,<sup>23</sup> and 41% of those do so multiple times a day.<sup>24</sup> This increasing penetration of online social networks into the fabric of the typical American's life provides fascinating, if challenging, opportunities for intervention design. Interventions delivered in the context of an online social network can leverage the availability of an individuals' self-identified social ties not only to optimise support for cessation, but also for active and passive distribution of the intervention through an individual's network to other smokers and beyond.

The importance of social networks on smoking behaviour and 'viral diffusion' has been seen in real-world networks. Data from the Framingham Heart Study demonstrated that smokers tend to cluster within their social networks and are significantly less integrated, that these patterns persist over time, and that clusters of smokers tend to quit together.<sup>25</sup> These findings suggest that interventions should target not just individual smokers but also their surrounding social network. This proposition is supported by robust evidence that social networks strongly affect social norms, and that within a network, norms may be altered by a single individual and perpetuated by other network members.<sup>26</sup> In addition, non-smokers may serve as 'weak ties' spanning clusters of smokers, analogous to a disease carrier who transports the disease between remote villages. Recruiting non-smokers to a cessation intervention may augment available social support for cessation, and increase social pressure (complex contagion effects) on other smokers to participate, thereby facilitating viral spread.<sup>27</sup>

This study aims to identify the variables that drive adoption and dissemination of an intervention throughout a network. The study examines this question within Facebook, the single largest and fastest growing online social network. As of 2013, approximately 143 million Americans use Facebook daily, with users globally sharing an average of 4.7 billion items of content daily.<sup>28</sup> Facebook enables individuals to create a profile, identify other members who are friends, exchange messages through multiple channels and—most relevant to this study—install small applications ('apps') created by third parties. These apps rely on 'viral' diffusion to grow their user base, and achieve this by inducing users to



Figure 1 Facebook Timeline.

'invite' their friends and enabling them to post information about the app to their personal 'Timeline' (essentially synonymous with the terms 'wall' or 'page') where it can be seen by others (figure 1). Data from Facebook suggest that individuals actively communicate with only 5% of their average 120 friends, but are passively exposed to information about 2–2.5 times as many.<sup>29</sup> By exposing a smoker's entire social network to a stream of 'pushed' information in real-time about their cessation progress (eg, 'Mary set a quit date'), it may be possible to significantly enhance social support for cessation (generating the response 'way to go Mary!' by network ties) and facilitate active and passive diffusion of the intervention.

The primary outcome metric of this study is the efficiency of this diffusion process, defined as the reproductive rate ( $R$ ). Online social networks depend on viral spread for dissemination of applications, a concept similar to snowball recruitment where investigators recruit only the 'seed' individual, and successive generations are recruited by the seed and their descendants. Within epidemiology,  $R$  is quantified as the mean number of secondary cases ('infections') that occur for a given 'infected' individual.<sup>30</sup> For online interventions, we can quantify the number of contacts of an individual and the duration that they are 'infectious' (ie, actively using an application). In this context,  $R$  can be expressed as:

$$R = t\beta Z$$

$t$  indicates the duration of being contagious (ie, the duration an individual uses an application),  $\beta$  is a constant of probability that determines the likelihood of spread from one individual to another for a given unit of time (referred to as 'contagiousness' hereafter), and  $Z$  is the number of contacts within the network. For an application with no diffusion,  $R$  will equal 0. For applications with  $R$  greater than 1, exponential growth will occur as each participant recruits/infects at least one other person; applications with  $R < 1$  will require ongoing seeding to maintain population growth. Increasing the amount of time ( $t$ ) that an application is used will increase the likelihood that it

spreads to new hosts. The goal of app developers is to reach an epidemic threshold where  $R$  exceeds 1 (ie, the app 'goes viral') and the application propagates autonomously, thus no longer requiring expenditures to recruit seed users.<sup>27</sup> While exceeding an  $R$  value of 1 is highly desirable, those cases where an epidemic threshold is not crossed can still serve as a multiplier for recruitment efforts. For example, 1000 individuals recruited to an intervention with  $R=0.2$  will yield 1250 participants over five generations of viral diffusion.

## Aims

The primary aim of the study is to identify and characterise the intervention characteristics that catalyse its diffusion through an online social network. We have tied our exploration to the concept of the  $R$  and the three independent variables that are its determinants—duration of use ( $t$ ), contagiousness ( $\beta$ ) and number of contacts ( $Z$ ). We empirically constructed domains of online intervention elements that we believed could impact each variable with minimal overlap: information content and proactive contact ( $t$ , duration of use), social communications ( $\beta$ , contagiousness) and non-smoker integration ( $z$ , number of contacts). We hypothesise that intervention variants containing greater information content, ongoing proactive contact and active communication strategies will outperform control application variants and have higher  $R$ s, and that their combination will be synergistic and will display positive interaction effects. We also hypothesise that the involvement of non-smokers as epidemiological 'carriers' will allow the application to spread more efficiently by bridging clusters of smokers.

The secondary aim is to identify and characterise the local (ego) networks of participants that effect diffusion and quitting behaviour. The characteristics of interest include smoking status, nicotine dependence, age, gender and local network characteristics (number of friends, network density and social position) as well as their number of friends already using the application. We hypothesise that local social network characteristics will predict the adoption behaviour of friends after invitation (active communication) or exposure (passive communication) from the participant. In addition, we hypothesise that invitation, adoption, utilisation and early cessation behaviour will display a complex contagion pattern, where increasing levels of network penetration (more friends that are pre-existing users of the application) will be associated with higher rates of diffusion, use of the application (duration and content exposure) and cessation behaviours (eg, setting of quit dates).

## METHODS/DESIGN

### Study design

This study involves two phases. Phase I was conducted from May 2012 to December 2012, and consisted of formative research designed to develop, test and optimise

multiple features of a Facebook application titled 'UbiQUITous'. Each feature is hypothesised to have a differential effect on the  $R$  of the application. Phase II is an ongoing randomised controlled trial that uses a fractional factorial design to determine the primary components of an online intervention for smoking cessation that determine its diffusion through a social network. The trial in phase II uses the UbiQUITous app as the study environment.

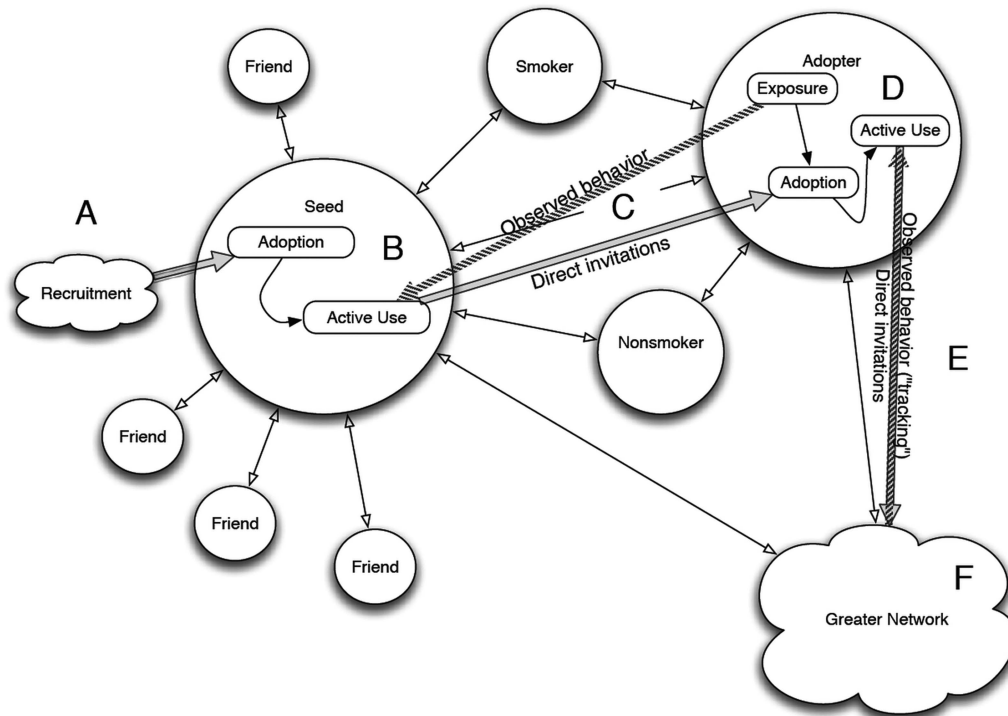
The generalised diffusion model guiding this study is presented in figure 2. Initial seed users are recruited (A) using purchased advertising within Facebook and earned media (unpaid publicity, such as a newspaper article or word of mouth). Data on resulting adoption (ie, app install) and utilisation (B and D, respectively)—including frequency and duration of use and content exposure—are automatically recorded for analysis. To evaluate diffusion (C and E), we use metrics of viral spread including the number of contacts ('friends') per user, the period of active use of the component ('infectious period') and the number of transmissions to determine the basic  $R$  of each studied component. Our software records all data in real-time to a relational database for later reconstruction of network maps and diffusion pathways. Detailed data collection methods are presented below. Based on the basic design patterns within Facebook, we divided  $\beta$  (the metric of contagiousness) into two forms:  $\beta$ -passive which results from observed behaviour and  $\beta$ -active which results from direct invitations or proactive, intentional contact from one user to another.

### Phase I: initial development and optimisation of the cessation application

In phase I of this trial, Collins'<sup>31</sup> MOST design method guided our efforts to break the proposed intervention into individual features and prototype each feature prior to evaluating the full intervention in a large-scale randomised trial. We augmented our internal software development team with expertise from an external graphic design firm to develop the visual components of the intervention. We prototyped multiple features, settling on six that were technically feasible, yielded useful data and resonated with our pilot users (see table 1). These six features could be set at multiple levels, each targeting a single element of diffusion ( $t$ ,  $\beta$  or  $Z$ ). To keep the size of the factorial model reasonable, we limited each feature to two levels (generally either on/off or high/low).

After development but prior to embarking on the full randomised trial, we consumer tested and refined each feature by testing a successive series  $\beta$  versions of the app. While this phase enabled us to detect programming and data recording errors, the primary focus was on iteratively evaluating and optimising the performance of each individual feature prior to the expense of a full randomised trial.<sup>31</sup> Facebook offers a free-development environment in which any third-party developer can create apps. For each  $\beta$  app, we launched the application





**Figure 2** Viral diffusion model.

features within Facebook and used paid advertising to recruit users. Based on user behaviour in the app and qualitative feedback gathered via short surveys to users, we made data-driven refinements in layout, presentation, content and message schedule, and evaluated their impact on the target metrics ( $R$ ,  $t$ ,  $\beta$  or  $Z$ ). Following refinement and optimisation, we proceeded to full recruitment and randomisation using a factorial model in phase II, described below.

#### Phase II: evaluation of diffusion in a large-scale randomised trial

The six features developed in phase I were translated into a factorial model. A full six-feature factorial would result in 64 separate cells. We simplified the matrix by combining features targeting the same variables to create four separate factors:  $t$  (expanded content and proactive contact),  $\beta$ -active (active diffusion: *invites* and social comparison) and  $\beta$ -passive (passive diffusion: *sharing*) and  $z$  (non-smoker supporters), resulting in a 16-cell factorial matrix (2 levels  $\times$  4 factors=16 cells). A final simplification eliminated the four cells that had no theoretical potential for diffusion, where  $\beta$ -active (*invites*) and  $\beta$ -passive (*sharing*) were disabled, resulting in a fractional factorial model with 12 cells (table 2).

#### Setting and participants

The randomised trial is conducted entirely within Facebook with all recruitment, screening, enrolment and randomisation automated by our clinical trials management software. Participants are registered users of

Facebook, a free social networking website. To be eligible, individuals must be current smokers, age 18 years or older and have an existing Facebook account. While seed enrolment targets English-speaking US residents, there are no language or residency restrictions. A 10% subsample is randomly selected from the initial seeds for additional data collection and follow-up.

#### Recruitment

Initial adopters ('seeds') are recruited primarily via online advertisements within Facebook that feature the app name, an app-related image and a short snippet of text advertising our free quit smoking app. Individuals clicking through to the app are shown a Facebook dialogue box asking for permission to install the application. Following app install, users provide informed consent for the study. Additional waves of participants are recruited via snowball methodology (ie, viral spread) as they are informed or invited to participate by friends within the network. Individuals who have a friend already using the application ('descendants') are enrolled in the study and represent the outcome of interest (ie, diffusion).

Inclusion criteria for *seed participants* are: US residency, current smoking, age 18 or older, have an English-language Facebook account and an email address, acceptance of Facebook permissions for app installation and provide study-informed consent. The only exclusion criterion is having one or more Facebook friends who have already installed the application. Age, existing friends who are already app users and location-

**Table 1** Application feature matrix

Target variable	Application feature	Levels
t	Expanded content (quit guides) <ul style="list-style-type: none"> <li>▶ A single generic quit guide supplemented by 9 additional topic-specific quit guides</li> <li>▶ 'Crave button' that when pushed randomly displays a humorous video from a content library</li> </ul>	On: has all quit guides, all crave content, has proactive app notifications, and gets an email on installation Off: has one quit guide, limited crave content, no proactive app notifications and no email on installation
$\beta$ -passive ( $\beta_p$ )	Proactive contact (app requests) <ul style="list-style-type: none"> <li>▶ Facebook requests: small icon appears on bookmark with flyover messages</li> <li>▶ App notifications: user receives a notification when friends achieve specific cessation milestones</li> <li>▶ Email: direct email on installation to the participant reminding them to come back</li> </ul>	
	Passive diffusion (sharing) <ul style="list-style-type: none"> <li>▶ Content (quit guides, crave content, check-ins, badges, money/life saved) can be shared on a user's Timeline</li> <li>▶ Badges earned and quit status are automatically posted to a user's Timeline</li> </ul>	On: content is sharable, Timeline posts are generated Off: content not sharable, Timeline posts not generated
$\beta$ -active ( $\beta_a$ )	Active diffusion (invites) <ul style="list-style-type: none"> <li>▶ During onboarding process, the user is prompted to identify and invite friends to either support them or quit with them</li> <li>▶ A persistent clickable interface element is available to invite friends</li> <li>▶ Users can post to a friend's Timeline using the 'Cure' and 'Capture' buttons (even if the friend has not installed)</li> </ul>	On: user can invite friends, post to friends' Timelines, and has both tabs of the leaderboard Off: user cannot invite friends, cannot post to friends' Timelines through the app, and only has the individual tab of leaderboard
	Social comparison (leaderboard) <ul style="list-style-type: none"> <li>▶ Compares individuals to others on various metrics. <a href="#">Table 1</a> compares individuals based on points earned in the app; <a href="#">Table 2</a> presents cumulative quitting metrics (\$ saved, life saved) for the participants' friends versus other participants' local networks</li> </ul>	
Z	Version for non-smoker supporters <ul style="list-style-type: none"> <li>▶ Non-smokers can install the application</li> <li>▶ Original quit guide tailored non-smokers; access to all other quit guides</li> <li>▶ Daily check-ins providing content on how to help a friend stay smoke-free</li> <li>▶ Otherwise identical experience to smokers</li> </ul>	On: smokers can have non-smoker supporters use the app Off: non-smoker supporters cannot use app (can install, but once declared non-smoker the app has no functionality)

related eligibility are assessed in real-time immediately upon installation. Informed consent is required in order to proceed into the app. Smoking status is assessed immediately after informed consent. Ineligible users who provide informed consent may still use the app, but are excluded from the study.

Subsample participants are randomly selected at a variable rate which is manually adjusted based on completion rates of the subsample survey to yield a final proportion of 10% of seed users. Subsample participants are reimbursed US\$20 per completed survey.

### Randomisation

*Seed users* are randomised to 1 of 12 cells using an adaptive 'biased-coin' strategy<sup>32</sup> which keeps the 12 cells in relative balance over the course of the trial. The probability of an individual being assigned to any given cell is adjusted in real-time by the clinical trials management system based on any pre-existing imbalance between the cells.

*Descendants* are users who have one or more Facebook friends and who have already installed the application ('parent'). Descendants install the app and accept

**Table 2** Cell manipulations

Cell	t	Z	$\beta$ -Active	$\beta$ -Passive	Diffusion three-level
1	0	0	0	0	X
2	0	0	0	1	0
3	0	0	1	0	1
4	0	0	1	1	2
5	0	1	0	0	X
6	0	1	0	1	0
7	0	1	1	0	1
8	0	1	1	1	2
9	1	0	0	0	X
10	1	0	0	1	0
11	1	0	1	0	1
12	1	0	1	1	2
13	1	1	0	0	X
14	1	1	0	1	0
15	1	1	1	0	1
16	1	1	1	1	2

Highlighted cells have  $\beta$ -active and  $\beta$ -passive disabled and are suppressed to create a 12-cell fractional factorial model.

informed consent in the same manner as seed users. They are assigned to the same cell as their parent. Descendants who have more than one friend in the study and for whom the diffusion channel (eg, active invite, Facebook ad) is unclear are assigned to the same cell as the friend who installed the app most recently. The designation of seed or descendant does not affect the user's app experience; they are simply identified as such in the relational database. Descendants may be smokers or non-smokers.

*Non-smokers* who do not have a friend in the app (ie, no parent seed) are assigned to the cell with all features enabled, but neither they nor their descendants are included in the study itself.

Facebook provides information on how an individual located the application, and if any of their friends are already users. Our application tracks potential paths of diffusion by embedding tracking tags within all links. New users who reach the app through an existing seed are identified in real-time and excluded from becoming seeds themselves.

### Intervention

The intervention is derived from the US Public Health Service (PHS) '5As' model (Ask, Advise, Assess, Assist and Arrange).<sup>2</sup> Content is based largely on PHS cessation materials for smokers supplemented by content written by the intervention team. Content is designed to motivate smokers to quit, provide support around a quit date, inform users of the benefits of quitting and build self-efficacy. On installation, users are greeted by the application's central character, Dr Youkwitz, who *Asks* participants if they smoke and *Advises* smokers to quit. He then *Assesses* their readiness to make a quit attempt and *Assists* them by providing a tool ('Quit Date Wizard') for planning a quit attempt and setting a quit date (see figure 3). Quit dates

are stored for analysis and are used for tailoring and targeting in other intervention components. If a user sets a quit date, the app displays a countdown to that date or an estimate of savings since that date (money saved, estimate of life saved). Users who do not set a quit date in their first visit may set one at any time. The application also *Arranges* follow-up in the form of daily check-ins with Dr Youkwitz who provide tailored and personalised information and support, and gather self-reported smoking status. Users randomised to cells that have the variable t turned on receive proactive Facebook app requests alerting them that a check-in is ready for them in the app. Participants who set a quit date are prompted at each check-in to confirm their quit date or update their smoking status. Smokers who have not set a quit date receive a variety of daily check-ins that include prompts to set a quit date, as well as evidence-based content incorporating the '5 Rs' (Relevance, Risk, Rewards, Roadblocks and Repetition) derived from the PHS guidelines. Users can receive check-ins for a year after their quit date.

The app employs simple game mechanics (points and badges) and a cartoon representation of Dr Youkwitz's lab, where the participant is exposed to smoking cessation information and tools (see figure 4). The longer a user stays engaged with the app, the more he/she is exposed to an unfolding narrative: Dr Youkwitz's experiments with a new anticraving drug have gone awry and have turned the user's friends into 'craving zombies'. Users can earn doses of a 'cure' by using various features of the app or by bringing their friends to the lab (ie, inviting to the app) to be cured and to provide support. This integration of game mechanics was designed to mirror the existing applications on Facebook, such as Farmville or Words With Friends.

This study utilises a factorial design to test the effects of multiple components of the cessation application (see table 1).

### t (Duration of exposure) features

Duration of exposure is maximised through expanded content and proactive contact. The app provides information in the form of a general quit guide or multiple topic-specific quit guides (eg, cessation and weight loss/maintenance, cessation and stress management) as well as library of short YouTube videos and animated gifs that can be accessed by pushing a 'Crave Button'. We hypothesise that the availability of smoking cessation informational content will be a strong driver of ongoing utilisation (thus increasing t). Proactive contact is implemented by encouraging users to come back to the app through a reminder that appears within the Facebook interface whether the user is using the application or not. These reminders also appear when a friend has installed the app, or when an installed friend hits a quit milestone.

### $\beta$ (Contagiousness) features

Individuals in online networks are exposed to personal information of approximately twice as many individuals





Figure 3 Application Quit Date Wizard.

as they nominate as buddies or with whom they actively communicate.<sup>20-29</sup> The app leverages this network phenomenon by manipulating sharing and competition-driven app use to drive contagion, both of which map

well to social support mechanisms that rely on information transfer and normative influence.

$\beta$ -Passive allows users to share app content (eg, when they set a quit date or earn a badge or content from a

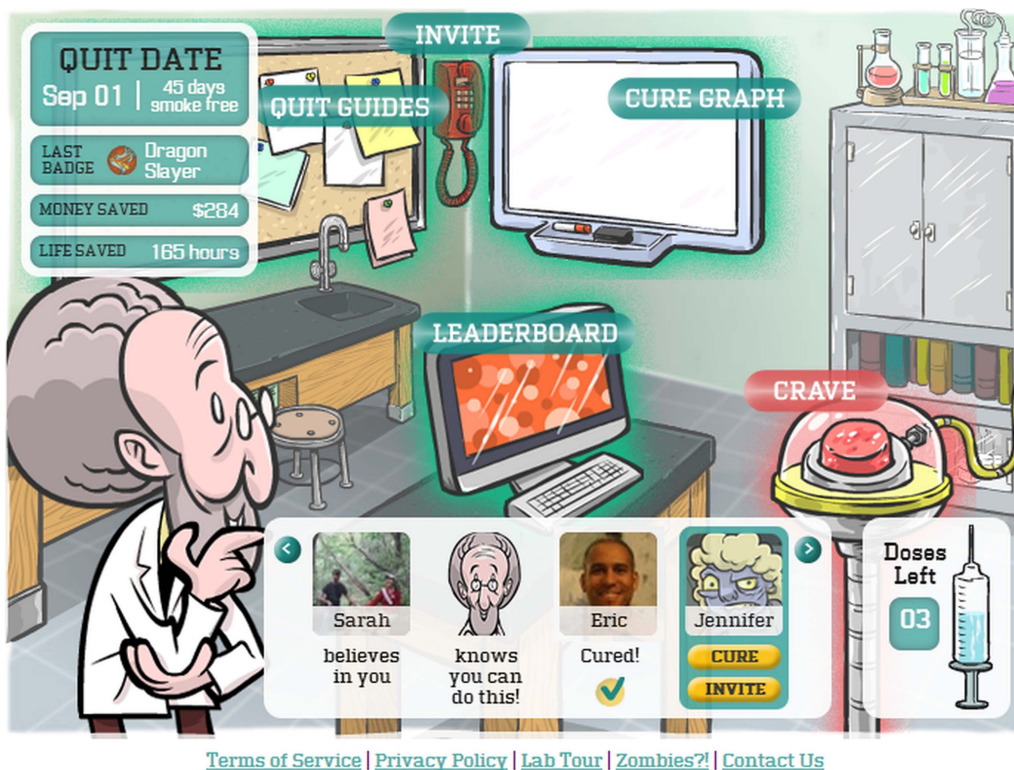


Figure 4 Application main screen.



quit guide) on their Timeline for friends to see. The app automatically posts on the participant's behalf when quit milestones are achieved (eg, setting a quit date, staying quit for consecutive days, reaching 1-month smoke-free). Each post to a participant's Timeline—either by themselves or by the app—generates opportunities for their friends to actively engage with the user's quit attempt by liking, commenting on, sharing the application-generated object or clicking on the shared content. Individuals who have not yet installed the application who click on app content are taken to a page with further information (eg, health benefits the user attained by reaching 1-month smoke-free) and encouraged to install the app to support their friend.

$\beta$ -Active allows participants to invite members of their Facebook network to install the app. The app encourages participants to invite others for cessation support and also to achieve game-based rewards. Participants may also share content from the application directly to a Facebook friend's Timeline or Wall. Network-level data are also displayed so that participants are exposed to goal-driven and normative information that compares them with others and to prespecified metrics (eg, number of friends with application installed, individual 'game points' earned via engagement with the application and hitting cessation-based milestones and collective life saved by the participant and their installed friends). The information and presentation are designed to encourage individuals to actively recruit others to participate.

### Z (number of contacts) feature

In order for an intervention targeted at smokers to spread with maximum efficiency from *cluster* to *cluster* (bridging) it needs to induce a 'carrier' state in non-smokers. A version of the app for non-smoker supporters allows non-smokers to provide support and has content tailored for non-smokers. Seed users are randomised to a version that can be shared with non-smoking friends or a version that is restricted to sharing with other smokers.

### Data collection and measures

The majority of data collection occurs through an application programming interface (API) provided by Facebook. The API allows our systems to interact directly with Facebook's database to retrieve data about individual users and their immediate social network. Since this study is a test of diffusion, we deliberately chose not to insert additional questions into the standard application installation process. Each participant is identified with a unique numeric identifier provided by Facebook.

To supplement the limited demographic data available through Facebook, we subsample 10% of seed users to further characterise study participants and to provide an estimate of intervention effectiveness. Measures collected from all seed (and descendent) users are listed

below. Measures collected only from the subsample are indicated as such.

### Facebook data

Data available from Facebook include email address, date of birth, gender, location, hometown, photos, likes, groups and a list of friends (including friends' birthdate, gender, location, likes, relationship to user and photos). Location and hometown information is optional within Facebook and not always available. Connections between a participant's friends are gathered automatically when available. Photos, groups, likes and location are used to construct and weight a social network graph using multiple co-occurrences as evidence of a stronger tie.

### Automated process tracking measures

A Facebook member may choose to install an application ('become infected') based on advertising or other earned media, observation of others' behaviour or direct invitation. Data on daily advertising expenditures, exposures and subsequent click-throughs are recorded automatically into the relational database. Standardised mechanisms within the Facebook API are used to record 'invitations', Timeline posts and subsequent 'acceptance' or click-throughs by individuals, allowing a clear chain of diffusion to be established back to an initial adopter and a precise calculation of R at each degree of separation. For application adoptions not specially mapped to an individual user (eg, where an individual is exposed to information about multiple friends using the intervention, but who is not specifically invited), we record their friend who installed most recently as a separate 'guessed' parent.

While abstinence is not a primary outcome metric in this study, information on quit dates is used as a marker of smoking status. This information will be used during analysis to extrapolate smoking status at arbitrary time points to reconstruct social network structure. In our earlier work, we have found that quit dates occurring in the past is a useful proxy for smoking status in descriptive analyses.<sup>20 22</sup>

For all participants, we record application installation, each return visit, specific pages viewed, total duration of the visit and use of app tools (eg, Quit Date Wizard, daily check-ins). Additionally, we record application uninstallation or blocking and 'likes' and 'dislike' tags. All application errors are recorded (eg, failure to post to a feed) to a standardised error log, as is downtime of the application and of Facebook itself. This error data are used in real-time to adjust performance, and if needed will be controlled for in final analyses.

### Baseline and follow-up self-report data (subsample only)

Users selected for the subsample are presented with a survey when they indicate smoking status in the app. The survey is presented within the Facebook frame, and consists of demographic, smoking status and nicotine dependence, social support and network size questions.



Subsampled participants are contacted via email at 30 days to take a web-based follow-up survey using a subset of baseline measures; a relatively short-time frame was selected to maximise response rates which were expected to be low. Non-responders receive a reminder email at 33 days, and are contacted directly by research staff through a private Facebook message at 34 days.

### Smoking status and nicotine dependence

Self-identified smoking status is assessed among all participants at enrolment. Subsample participants also report readiness to quit<sup>33</sup> and nicotine dependence is measured with the 'time to first cigarette' item.<sup>34</sup>

### Social support for cessation

Subsample participants complete an adapted version of the Partner Interaction Questionnaire (PIQ).<sup>35</sup> This measure assesses receipt of specific positive and negative behaviors from an individual who has followed the participant's efforts to quit smoking most closely.

### Network size

Subsample participants complete a series of questions about "how many people named [first name] do you know."<sup>36</sup> The names selected satisfy a 'scaled-down condition' such that, for example, if 15% of the population is men between age 21 and 40, then 15% of the people asked about also must be men between age 21 and 40. We implemented the names inventory as in McCormick *et al*<sup>36</sup> and the Pew Internet Survey in 2011.<sup>37</sup>

### Power analysis

We have compensated for the difficulty of estimation of sample size in this field by leveraging our capacity for scalable, low-cost recruitment. As an application becomes more accepted and valued by a group, others are more likely to value it.<sup>38</sup> Since this process is unpredictable at the individual level, the end exponential effects are highly unpredictable. The common-sense approach to this, proposed by Watts and Dodds,<sup>39</sup> is 'the big seed model', and involves seeding the application to as many initial individuals as possible, rather than carefully targeted few. In a small network this can be an issue, as initial seeds may know each other, thus contaminating the diffusion metrics; however, in large social utilities with tens to hundreds of millions of users this is statistically less likely.

Examples of basic R for viral marketing campaigns in other online modalities have ranged from 0.041 to 2.<sup>39 40</sup> The study is powered for a sub 1.0 R and at only the first degree (R0) to guarantee productive analysis and results even if the intervention does not reach 'viral threshold'. Using data from prior studies in the business and social marketing literature, we estimated sample size calculations for individual cell comparisons and estimate R ranging from 0.1 to 0.5. We calculated that a study size of N=8000 would provide 88% power to examine all between-factor analyses with a minimal detectable difference in basic R of 0.1. Increasing the sample size to

12 560 yields the ability to examine the interaction effects at the same detectable difference and a power of 80%.

### Statistical analyses

Outcome data will be obtained in the form of counts of new cases arising from direct contact with primary seed subjects, and exposure in the count of total seed subjects under each treatment cell. The design was a fractional factorial; there was no manipulation without passive and active diffusion (see table 2 for manipulations). The diffusion manipulations will be collapsed into a single three-level categorical treatment. Analysis will be conducted using Poisson regression, using Generalised Linear Models with a log-link and Poisson family. Design variables will be entered as predictors representing the treatment combinations, with the reference category representing minimal content, no non-smoker support and only passive diffusion. Results will be reported as baseline R for the reference group, and incidence rate ratios for all other entries and hypotheses were tested at  $\alpha=0.05$ . We will test interactions for entry into the model, and omit the interactions if they are not significant. Post hoc comparisons may be made after fitting the regression model using the Wad test.

## ETHICS AND DISSEMINATION

### Informed consent

We use a two-step consent process: the participant first provides consent to Facebook for the release of their data via a dialogue box within Facebook itself (see figure 5), and then provides informed consent to the study. This study was deemed eligible for abbreviated consent as per Federal regulations.

### Dissemination

In theory, study data can be anonymised and made suitable for data sharing; however, this has proved highly challenging in practice and risks of disclosure remain in many datasets.<sup>41</sup> If anonymity of participants can be assured, we intend to make data available to other investigators through either a National Institutes of Health (NIH) mechanism (CaBIG) or a non-profit academic mechanism such as the Dataverse project. We intend to repackage our source code as an open source platform for performing research within Facebook, and welcome potential collaborators.

The study results will be disseminated through conference presentations and peer-reviewed manuscripts.



Provide **UbiQUITous** your public profile, friend list, email address, relationships, birthday, groups, hometown, current city, photos, likes and your friends' relationships, birthdays, groups, current cities, photos and likes.

[App Terms](#) · [Privacy Policy](#)

**Figure 5** Facebook data transfer consent screen.



Initial results of advertising and recruitment methods have been presented in abstract form at academic conferences, while implementation and programming methods have been presented by the development team at engineering conferences. The main outcomes of the trial, planned social network analyses and secondary data exploration will be presented at future conferences and published in the peer-reviewed literature. Given the topic of this project, however, we are equally interested in novel forms of distribution of the findings themselves through social networks. We are experimenting with building audiences with Tumblr (a blogging platform) and Twitter, and intend to publish at least a portion of the results in open access journals.

## DISCUSSION

This protocol describes an experiment to explore a novel mechanism to disseminate evidence-based treatment for smoking cessation using one of the largest online social networks, Facebook. Results from this study will add to the knowledge base about constructing interventions capable of self-propagation and distribution and how they may influence behaviour in local networks. Interventions delivered through online social networks offer potential not only to enhance social support but also to enrich social influence. In an existing network, an individual who quits smoking exerts an effect on network ties, causing collateral or even cascading smoking cessation across multiple degrees of separation and potentially producing a cumulative impact greater than that would be predicted by efficacy rates alone.<sup>25</sup> This cascade has the potential to serve as a profound multiplier for public health spending.<sup>42</sup> We anticipate that a new generation of research protocols leveraging complexity science will explore not just viral diffusion, but the interdependent impact of diffusion and uptake on social and behavioural processes.

There are several limitations to this study that stem from the nature of Facebook itself. The most significant is the trade-off between maximising dissemination and the collection of personal information. Additional data collection would have been 'invasive' relative to consumer expectations within online social network and would potentially suppress the primary outcome of the trial (ie, viral spread). We deliberately chose not to evaluate cessation outcomes in this trial since doing so would have added a significant burden to participants and dampened our outcome of interest. Future research should address the question of efficacy. At the time of writing the initial proposal for this project, we acknowledged a risk that prior to, or during this study, the social network landscape might change or Facebook could change its internal mechanisms. We based this proposal (and the pilot work) on what appear to be the basic and common elements of the platform that seemed unlikely to change. Not surprisingly, a number of minor Facebook platform changes occurred prior to the phase

II recruitment, requiring protocol changes that are reflected in this document. We have found having an active, in-house engineering team invaluable in keeping up with changes in the Facebook environment.

Ultimately, if this intervention approach succeeds in demonstrating viral spread, the project will have the potential to substantially shift how tobacco treatment, or any other health behaviour, services are marketed and delivered. Viral distribution of a behavioral intervention through existing social networks could be applied to multiple health conditions, including smoking, obesity, nutrition and alcohol. Our hope is that this study informs a near-term, future generation of effective health interventions to be disseminated to large populations in a low-cost, efficient manner.

## Author affiliations

<sup>1</sup>Division of Pulmonary and Critical Care, Georgetown University Medical Center, Washington DC, USA

<sup>2</sup>Department of Health, Behavior and Society, The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

<sup>3</sup>Schroeder Institute for Tobacco Research and Policy Studies, American Legacy Foundation, Washington DC, USA

<sup>4</sup>North American Research and Analysis, Inc, Faribault, Minnesota, USA

<sup>5</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

<sup>6</sup>Department of Oncology, Georgetown University Medical Center/Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, USA

**Contributors** NKC and ALG participated in study concept. NKC, EPW, ALG participated in study design. NKC, MAJ and ALG participated in acquisition of data. NKC, EPW and ALG participated in statistical analysis. NKC, MAJ, JS, EPW and ALG participated in drafting of the manuscript. NKC, MAJ, JS, EPW, and ALG made comments on the manuscript. NKC, EPW, and ALG participated in obtaining funding.

**Funding** Primary funding for this work was from the National Cancer Institute at the National Institutes of Health (1R01CA155369).

**Competing interests** NKC is an employee of MeYou Health LLC, whose parent company's product line includes an online tobacco cessation intervention. ALG and MAJ are employees of Legacy, a non-profit public health foundation that runs BecomeAnEX.org, an online tobacco cessation intervention.

**Patient consent** Obtained.

**Ethics approval** The study protocol was approved by Schulman Associates Institutional Review Board (IRB; formerly Independent IRB).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

## REFERENCES

- Centers for Disease Control and Prevention (CDC). Current cigarette smoking among adults—United States, 2011. *MMWR Morb Mortal Wkly Rep* 2012;61:889–94.
- Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services (USDHHS); 2008. 257 p.

3. Kokkinides VE, Ward E, Jemal A, *et al.* Under-use of smoking-cessation treatments: results from the National Health Interview Survey, 2000. *Am J Prev Med* 2005;28:119–22.
4. Orleans CT. Increasing the demand for and use of effective smoking-cessation treatments reaping the full health benefits of tobacco-control science and policy gains—in our lifetime. *Am J Prev Med* 2007;33(Suppl 6):S340–8.
5. Shiffman S, Brockwell SE, Pillitteri JL, *et al.* Use of smoking-cessation treatments in the United States. *Am J Prev Med* 2008;34:102–11.
6. Abrams DB, Orleans CT, Niaura RS, *et al.* Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: a combined stepped-care and matching model. *Ann Behav Med* 1996;18:290–304.
7. Who's online: Internet user demographics. Pew internet & American life project 19 May 2013: [http://pewinternet.org/Static-Pages/Trend-Data-\(Adults\)/Whos-Online.aspx](http://pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx) (accessed 3 Sep 2013).
8. Fox S. *Online health search 2006*. Washington, DC: Pew Internet and American Life Project, 2006.
9. Cobb NK, Graham AL. Characterizing Internet searchers of smoking cessation information. *J Med Internet Res* 2006;8:e17.
10. Myung SK, McDonnell DD, Kazinets G, *et al.* Effects of web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. *Arch Intern Med* 2009;169:929–37.
11. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. *Addiction* 2009;104:1792–804.
12. Chen YF, Madan J, Welton N, *et al.* Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. *Health Technol Assess* 2012;16:1–205, iii–v.
13. Civljak M, Stead LF, Hartmann-Boyce J, *et al.* Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2013;(7):CD007078.
14. Muñoz RF, Lenert LL, Delucchi K, *et al.* Toward evidence-based Internet interventions: a Spanish/English web site for international smoking cessation trials. *Nicotine Tob Res* 2006;8:77–87.
15. Swartz LH, Noell JW, Schroeder SW, *et al.* A randomised control study of a fully automated internet based smoking cessation programme. *Tob Control* 2006;15:7–12.
16. Strecher VJ, McClure JB, Alexander GL, *et al.* Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med* 2008;34:373–81.
17. Graham AL, Cobb NK, Papandonatos GD, *et al.* A randomized trial of Internet and telephone treatment for smoking cessation. *Arch Intern Med* 2011;171:46–53.
18. Bock BC, Graham AL, Sciamanna C, *et al.* Smoking cessation treatment on the Internet: content, quality, and usability. *Nicotine Tob Res* 2004;6:207–19.
19. Bock BC, Graham AL, Whiteley JA, *et al.* A review of web-assisted tobacco interventions (WATIs). *J Med Internet Res* 2008;10:e39.
20. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. *Am J Public Health* 2010;100:1282–9.
21. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7:e11.
22. Cobb NK, Graham AL, Bock BC, *et al.* Initial evaluation of a real-world Internet smoking cessation system. *Nicotine Tob Res* 2005;7:207–16.
23. Duggan M, Brenner J. The demographics of social media Users. Washington, DC: Pew Internet & American Life Project, 2013, 14 p.
24. Rainie L, Smith A, Duggan M. *Coming and going on Facebook*. Washington, DC: Pew Internet and American Life Project, 2013, 7 p.
25. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008;358:2249–58.
26. Latkin C, Donnell D, Celentano DD, *et al.* Relationships between social norms, social network characteristics, and HIV risk behaviors in Thailand and the United States. *Health Psychol* 2009;28:323–9.
27. Cobb NK, Graham AL. Health behavior interventions in the age of facebook. *Am J Prev Med* 2012;43:571–2.
28. *Facebook's growth in the past year*. Facebook media set. (2013). May 2013 (accessed 3 Sep 2013).
29. The size of social networks: Primates on Facebook. *The Economist* 2009:13176775.
30. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface* 2005;2:281–93.
31. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007;32(Suppl 5):S112–18.
32. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:515–19.
33. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. *Am Psychol* 1992;47:1102–14.
34. Baker TB, Piper ME, McCarthy DE, *et al.* Time to first cigarette in the morning as an index of ability to quit smoking: implications for nicotine dependence. *Nicotine Tob Res* 2007;9(Suppl 4):S555–70.
35. Graham AL, Papandonatos GD, Bock BC, *et al.* Internet-vs. telephone-administered questionnaires in a randomized trial of smoking cessation. *Nicotine Tob Res* 2006;8(Suppl 1):S49.
36. McCormick TH, Salganik MJ, Zheng T. How many people do you know?: efficiently estimating personal network size. *J Am Stat Assoc* 2010;105:59–70.
37. Hampton K, Goulet LS, Rainie L, *et al.* Social networking sites and our lives. Washington, DC: Pew Internet & American Life Project, 2011.
38. Salganik MJ, Dodds PS, Watts DJ. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 2006;311:854–6.
39. Watts DJ, Dodds PS. Influentials, networks, and public opinion formation. *J Consum Res* 2007;34:441–58.
40. Kalyanam K, McIntyre S, Masonis JT. Adaptive experimentation in interactive marketing: the case of viral marketing at Plaxo. *J Interact Mark* 2007;21:72–85.
41. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Rev* 2010;57:1701.
42. Fowler JH, Christakis NA. Estimating peer effects on health in social networks: a response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. *J Health Econ* 2008;27:1400–5.