

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([see an example](#)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diffusion of an Evidence-Based Smoking Cessation Intervention Through Facebook: A Randomized Controlled Trial Study Protocol
AUTHORS	Cobb, Nathan; Jacobs, Megan; Saul, Jessie; Wileyto, Paul; Graham, Amanda

VERSION 1 - REVIEW

REVIEWER	Lindsay Stead University of Oxford, UK
REVIEW RETURNED	17-Oct-2013

GENERAL COMMENTS	No further comments. This is a clear description of a study protocol. I will look for the results with interest.
-------------------------	--

REVIEWER	Diana D. McDonnell University of California, Berkeley, USA
REVIEW RETURNED	21-Oct-2013

GENERAL COMMENTS	This is a discussion of a study protocol, not review of study results, so some of the above do not apply. #4: I commented where I think they need to expand this, it's not often, but not never.
-------------------------	---

- Reviewer 2 also provide a marked copy which is available upon request from the Publisher.

VERSION 1 – AUTHOR RESPONSE

1. Page 2, Introduction paragraph: “Diffusion can be measured using R...” – I would strengthen this statement -- not only CAN diffusion be measured this way, but it often is, e.g., infectious diseases. This is obviously just the abstract, so it has to be brief, but it's important to let the reader know you're using a well known / used measure.

RESPONSE: We have rephrased the sentence to state that ‘Diffusion is frequently measured using R...’ (Pg 2 – Abstract)

2. Page 3, paragraph 1: “Smoking remains the leading cause of 443,000 preventable deaths and \$200 billion in excess cost in the U.S. each year, making large scale reduction in smoking prevalence a public health imperative” – This statement needs a citation. If it's the same as one of the below, I'd still put it here too, as this is a very specific set of numbers.

RESPONSE: We have added the appropriate references. (Pg 3, para 1)

3. Page 3, paragraph 2: “The majority (85%) of US adults are Internet users, including populations at disproportionate risk for smoking such as African Americans (85%) and those with incomes less than \$30,000/year (76%)[6] and 6-9% of all Internet users (> 10 million adults) search for quitting smoking annually.” – Start a new sentence here, as all the % get confusing and this is an interesting / important statement all by itself.

RESPONSE: We have made the requested change. (Pg 3, para 2)

4. Page 3, paragraph 2: Clarify whether these parenthetical numbers are the proportion of this group with internet access, the proportion who smoke, or some combination thereof.

RESPONSE: We have clarified this in the text. (Pg 3, para 1)

5. Page 3, paragraph 2: “In such bounded networks...” – Explain more what this means.

RESPONSE: We have clarified this in the text by dropping the confusing word ‘bounded’ and being more explicit: “...participation is limited by affiliation with a particular behavior (e.g., quitting smoking).” (Pg 3, para 2)

6. Page 4, paragraph 2: “Recruiting non-smokers to a cessation intervention may...” – A very interesting -- and unusual -- proposition. Is there literature about something similar being used in other fields (i.e., dissemination of something beneficial and/or behavioral, by people not directly involved)?

RESPONSE: We are not aware of this approach being discussed or implemented in a way that mentions or takes advantage of formal network theory – despite the fact the approach derives from traditional concepts of carriers and herd immunity. We make no claim that it is particularly novel, but despite a thorough literature search were unable to find other researchers describing its application in health behavior. We have added a reference to one of our own papers (a commentary) that explores how this approach could be applied to two other published studies on Facebook. (Pg 4, para 2)

7. Page 4, paragraph 3: “These apps rely on “viral” diffusion to grow their user base, and achieve this by inducing users to “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.” – If you have enough space, it would be helpful to have a screenshot here of a generic Facebook page that includes an app, so the non-Facebook readers get a better sense of what you’re talking about.

RESPONSE: We have inserted an additional figure and requested the space from the editor. (Page 4, last para)

8. Page 5, paragraph 2: “The goal of app developers is to reach an epidemic threshold where R exceeds 1 (i.e., the app “goes viral”) and the application propagates autonomously, thus no longer requiring expenditures to recruit seed users.” – Very interesting application of infectious disease diffusion. Are there examples in the literature of this equation being used similarly, for other behavioral and/or recruitment efforts?

RESPONSE: The equation is commonly used by online application designers – the term ‘to go viral’ comes from the online world where videos, apps or games suddenly undergo massive network-based distribution. While there is a small amount of research looking at diffusion in large-scale networks as well as non-experimental/academic work noting the benefits of viral spread, we are unaware of any work in the peer-reviewed literature on health behavior explicitly using the reproductive rate as an

outcome metric. We have added a citation of one of our own papers calling for better use of epidemiologic methods in evaluating online health behavior applications (Cobb NK, Graham AL. Health behavior interventions in the age of Facebook. *Am J Prev Med* 2012;43(5):571-2). (Pg 5, para 2)

9. Very nice intro. One area I'd like you to address is how likely people trying to quit are to want to share this with their networks -- i.e., is there literature saying that people trying to quit use the act of telling others about their attempt as a way to gain support or, conversely, that they want to keep their attempt private because they're afraid that they will fail and therefore disappoint others? Maybe different kinds of people are more likely to do one or the other?

RESPONSE: We agree that there are likely subgroups of smokers who are more/less interested in sharing their quit attempts with members of their existing social network. Over the past decade our group has developed multiple online, social systems that allow smokers to interact with others to marshal social support and enhance self-efficacy (references to this work are throughout the manuscript). These systems, and those developed and evaluated by others, usually connect individuals with strangers in a semi-anonymous fashion – similar to Alcoholics Anonymous. We are unaware of literature looking specifically at anonymity and participants' desire to share or hide their quit attempts from their existing social networks. To begin to address this question without unduly burdening participants with lengthy assessment instruments, we included a brief measure of perceived support for cessation called the Partner Interaction Questionnaire. This instrument measures receipt of positive and negative support from someone who follows their efforts to quit smoking. We will examine this measure in relation to use of the app to begin to address this question. We have added language on Page 14 that explains the inclusion of these measures. (Pg 14, para 3)

10. Page 6, paragraph 4: Define “earned media”

RESPONSE: We have clarified what we mean by earned media: “unpaid publicity, such as a newspaper article or word of mouth.” (Pg 6, final para)

11. Page 7, paragraph 1: “All data are recorded in real-time to a relational database for later reconstruction of network maps and diffusion pathways.” – Is this something that Facebook does for all apps, or is this something you created yourself (or a combination thereof)?

RESPONSE: All data, whether received from Facebook, or generated by our own software is recorded by our software into a database under our control. We have clarified this on Page 7 of the manuscript: “Our software records all data in real-time into a relational database ...” (Pg 7, para 1)

12. Page 7, paragraph 3: “After development but prior to embarking on the full randomized trial, we consumer tested and refined each feature.” – Say something -- at least briefly -- about how you (and who) developed the app in the first place.

RESPONSE: We have added a sentence about our internal software development team and external consultants. (Pg 7, para 3)

13. Page 7, paragraph 3: “We launched the application features within Facebook...” – Say a bit more about this -- can anyone do it, does it cost money, what type of format was it in. Or give a reference for people who want to do this and have never done it before.

RESPONSE: There are no licensing fees to Facebook, however the multidisciplinary skill set (software engineers, database specialists, visual designers and interventionists) required precludes most groups from developing their own software. Our internal software development team includes

industry experience and public health communications expertise. We augmented their skill set with graphic design consultants specifically to develop the visual components of the app. (Pg 7, para 3)

14. Page 8, paragraph 1: "...made data-driven refinements in layout, presentation..." – Based on what / whose data? Facebook's that you have access to or something you collected on your own. How did you know what was ideal or better vs. worse -- i.e., what about the data indicated that something needed to be changed? Did all this happen while the app was "live" on Facebook, or in some test situation?

RESPONSE: In this brief sentence we refer to the software development process of the application, using relatively standard development techniques for testing and refinement. At the bottom of Page 7, we have clarified that the data we collected on utilization to drive these changes occurred with beta versions of the app, deployed with real users in Facebook, but prior to the launch of the study. (Pg 7, final para)

15. Page 7, paragraph 3: "Following refinement and optimization, we proceeded to full recruitment and randomization using a factorial model in Phase II." – I know that this part is not the focus of this particular paper, but remember that part of sharing results is so that others can replicate your study / intervention, so include enough info / references here so that people who have not created an app before get at least a general sense of what is involved.

RESPONSE: Unfortunately, the software design and development process – similar to pharmaceutical drug or medical device development – is beyond the reasonable scope of a study protocol manuscript. We have written the protocol to match the grant proposal and IRB protocols for the study itself, but have not included our detailed work plan for the software and intervention development. We appreciate that readers may be interested in these details and fully intend to publish a manuscript on the topic in the future alongside our main outcomes paper. We hope the reviewer is comfortable with the limited description of the software engineering process for this study protocol paper.

16. Page 8, paragraph 1: Since this is pretty complicated and methodologically quite important, it would be helpful to have a table labeling the full 64 cells, and somehow highlighting the ones you combined or eliminated as well as the remaining 12.

RESPONSE: The full matrix was 16 cells (we consolidated factors to reach a 4x4 rather than 6x6 matrix), with 4 cells suppressed to yield fractional-factorial of 12. The matrix is presented later in the manuscript, but we have added a reference to it near the location noted by the reviewer. (Pg 8, para 1)

17. Page 8, Recruitment paragraph: What information does the initial app / ad give people? How / when is eligibility assessed?

RESPONSE: We have added additional information about the recruitment and screening process on Pages 8 and 9 (Recruitment).

18. I'm confused about how seed participants vs. descendants download the app and provide informed consent -- what's the difference and how do you differentiate them so that descendants do not present as seeds?

RESPONSE: The process of app installation and consent is identical for all users - the only difference is that 'descendants' by definition have a trail of data pointing back to another user as a 'parent'. Individuals that have data that indicate they were 'infected' by another individual are thus ineligible to

be enrolled as a seed or randomized, but serve as evidence of successful diffusion for their parent seed. We have clarified how this occurs in real time in the manuscript in the Randomization section on page 9. “Facebook provides information on how an individual located the application, and if any of their friends are already users. Our application tracks all potential paths of diffusion by embedding tracking tags within links. New users that reach the app through an existing seed are identified in real time and excluded from becoming seeds themselves.” (Pg 9, para 3)

19. Page 9, paragraph 2: “The probability of an individual being assigned to any given cell is adjusted based upon any pre-existing imbalance between cells.” – Based on what? And how does what- / whom- ever randomizes get this information so it can be used to adjust the randomization probabilities. Is this in real time?

RESPONSE: All aspects of recruitment, screening, and randomization occur in real-time. We have made this explicit with additional language in the ‘Settings and Participants’ section on Page 8 that states “The randomized trial is conducted entirely within Facebook with all recruitment, screening, enrollment and randomization automated by our clinical trials management software.” (Pg 8, para 2)

20. So are the non-smokers seeds or descendants or neither?

RESPONSE: Non-smokers can be descendants (counting towards R), but never seeds under the protocol. We have clarified this in the Randomization section. (Pg 9, para 3)

21. Page 10, paragraph 2: Is check in active (i.e., pushed to users) or passive (users do it on their own time / schedule)?

RESPONSE: Users are randomized to this precise variable condition – those in a t(+) condition receive proactive messaging to check-in, while those in a t(-) condition do so passively. We have clarified this in the manuscript. (Pg 10, para 1)

22. Page 10, paragraph 2: What determines someone's follow up length?

RESPONSE: The intervention will be available to all users for one year. However, data from our initial pilot work showed very little diffusion after the first few days of app use. Therefore, our main outcome analyses are tied to a 1 month follow-up. We have clarified the period of follow-up in the intervention section (Page 10) and in the analysis section. The last sentence of the first paragraph in the Intervention section reads “Users can receive check-ins for a year after their quit date.” (Pg 10, para 1)

23. Page 10, bottom of the page: “This fusion of game mechanics with cessation content was designed to mirror many existing applications on Facebook.” – Such as? Are any cessation or health related? Is there any evidence of efficacy? Any numbers related to users or engagement or dissemination?

RESPONSE: We have added several examples of Facebook apps in the text on Page 10. “This integration of game mechanics was designed to mirror existing applications on Facebook, such as Farmville or Words With Friends.” We did not intend to mirror cessation or health applications (which are very rare and seem to have very few users), but rather applications that are widely used and would be a point of reference and familiar to users. In other words, we wanted our app to look like other Facebook apps, not like a health intervention. (Pg 10, final para)

24. This reminds me, was there any incentive given to either the seed or descendant participants?

RESPONSE: We have clarified the use of incentives for the sub-sample in the Recruitment section on page 9. No incentives were used otherwise. (Pg 9, first para)

25. Page 13, Data Collection and Measures: "The majority of data collection occurs through an application programming interface (API) provided by Facebook." – I asked about this earlier and you cover it here, but I think a general statement earlier (even if it's to say that this will be covered later) would be helpful.

RESPONSE: We have made the requested change on Page 7 as follows: "Detailed data collection methods are presented below."

26. Page 13, bottom of the page: Is this FB data available for everyone or is it optional info people can give to FB, in which case it might be missing for some of your participants?

RESPONSE: The basic demographic information is required by Facebook, and we collect all of this data for all participants. As part of the informed consent process individuals provide their permission to Facebook for the app to retrieve the information. Users that decline to provide the information are unable to install the application or interact with the intervention or study system. We have clarified that user's locations are optional in Facebook and not always collected. (Pg 13, para 3)

27. It would be helpful to talk a little about how / if you think it will ever be possible to assess actual health outcomes / improvements through things like FB or if this will always have to rely on a separate mechanism. That is, is there a real way to tell if the apps are actually "working" vs. if they are just being used and disseminated?

RESPONSE: We believe that Facebook is simply a communications channel, like the phone, email, the web or text messaging. It is possible to use any of these channels to assess health outcomes, though attrition and self-report bias are inherent challenges. We have added brief acknowledgement of this tension, stating "We deliberately chose not to evaluate cessation outcomes in this trial since doing so would have added significant burden to participants and dampened our outcome of interest (diffusion). Future research should address the question of efficacy." (Pg 17, last paragraph)