HHS Public Access
Author manuscript
J Epidemiol Community Health. Author manuscript; available in PMC 2019 February 01.

Published in final edited form as:
J Epidemiol Community Health. 2018 February ; 72(2): 173–178. doi:10.1136/jech-2017-209976.

Generalizability of an Online Randomized Controlled Trial: an Empirical Analysis

Cheng Wang[1,2,3,4], **Katie R. Mollan**[5,*], **Michael G. Hudgens**[6], **Joseph D. Tucker**[2,3,4], **Heping Zheng**[2,3], **Weiming Tang**[2,3,4,*], and **Li Ling**[1,*]

[1]Faculty of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Dermatology Hospital of Southern Medical University, Guangzhou, China

[3]Guangdong Provincial Dermatology Hospital, Guangzhou, China

[4]University of North Carolina Project-China, Guangzhou, China

[5]Center for AIDS Research, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

[6]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Background—Investigators increasingly use online methods to recruit participants for randomized controlled trials (RCTs). However, the extent to which participants recruited online represent populations of interest is unknown. We evaluated how generalizable an online RCT sample is to men who have sex with men (MSM) in China.

Methods—Inverse probability of sampling weights (IPSW) and the G-formula were used to examine the generalizability of an online RCT using model-based approaches. Online RCT data and national cross-sectional study data from China were analyzed to illustrate the process of quantitatively assessing generalizability. The RCT (identifier NCT02248558) randomly assigned

*Li Ling, Katie Mollan and Weiming Tang are co-correspondence authors: Li Ling, Faculty of Medical Statistics and Epidemiology, School of Public Health, SunYat-sen University, Guangzhou, China. lingli@mail.sysu.edu.cn, Katie Mollan, Center for AIDS Research, School of Medicine, University of North Carolina, Chapel Hill, NC, USA. kmollan@email.unc.edu, Weiming Tang, Dermatology Hospital of Southern Medical University, Guangzhou, China, Guangdong Provincial Dermatology Hospital, Guangzhou, China, University of North Carolina Project-China, Guangzhou, China.

Supplementary Data
Supplementary data are available online

Contributorship Statement
1) Conception and design of study: Michael G. Hudgens, Joseph D. Tucker, Katie R. Mollan, Cheng Wang, and Ling Li; 2) Acquisition of data: Joseph D. Tucker, Weiming Tang, Cheng Wang, and Heping Zheng; 3) Analysis and/or interpretation of data: Cheng Wang, Katie R. Mollan, Michael G. Hudgens; 4) Drafting the manuscript: Cheng Wang. 5) Revising the manuscript: Katie R. Mollan, Joseph D. Tucker, Michael G. Hudgens, Weiming Tang, Ling Li, and Heping Zheng.

Competing Interests: None declared

participants to a crowdsourced or health marketing video for promotion of HIV testing. The primary outcome was self-reported HIV testing within 4 weeks, with a non-inferiority margin of −3%.

Results—In the original online RCT analysis, the estimated difference in proportions HIV tested between the two arms (crowdsourcing - health marketing) was 2.1% (95% confidence interval (CI): −5.4% to 9.7%). The hypothesis that the crowdsourced video was not inferior to the health marketing video to promote HIV testing was not demonstrated. The IPSW and G-formula estimated differences were −2.6% (95%CI: −14.2 to 8.9) and 2.7% (95%CI: −10.7 to 16.2), with both approaches also not establishing non-inferiority.

Conclusions—Conducting generalizability analysis of an online RCT is feasible. Examining the generalizability of online RCTs is an important step before an intervention is scaled-up.

Keywords

crowdsourcing; generalizability; G-formula; HIV testing; inverse probability weighting; IPSW; non-inferiority; online RCT

INTRODUCTION

Investigators increasingly use online methods to recruit participants for randomized controlled trials (RCTs).[1–3] Online RCTs recruit participants, allocate interventions, measure outcomes, or collect data partly or fully using the Internet.[3] Online RCTs are increasingly used because they can efficiently recruit large samples (especially from groups that are small),[4] make studies easier to replicate and have some automation processes which facilitate study implementation[5–8] compared to offline RCTs. However, such RCTs may have poor generalizability.[1,9,10] For instance, online studies often have lower retention rates than offline research.[11,12] Online research is also more likely to recruit educated, Internet-using populations[1,13] who may not be representative of the larger population of interest.

Statistical methods have been developed which provide formal, quantitative, theoretically justified inferential approaches for generalizing RCT findings to populations of interest. These approaches include nonparametric direct standardization[14,15] and model-based methods.[16–18] To date, these types of methods have not been employed to generalize results of online RCTs, which is critical for practitioners and policy makers considering scaling up the results based on online trials.[19]

Herein we examine the generalizability of an online RCT using inverse probability of sampling weights (IPSW)[17,18] and the G-formula.[20–23] In particular, we use online RCT data[24] and national cross-sectional study data from China[25] to illustrate the process of quantitatively assessing generalizability of results from an online RCT.

METHODS

Study population

The crowdsourced online RCT in China analyzed here has been described in detail elsewhere.[24,26] Briefly, in 2014, our study team applied the crowdsourcing method to generate a one-minute video for promoting HIV testing among Chinese men who have sex with men (MSM). Crowdsourcing is the process of shifting an individual task to a large group in the form of a contest or open call.[27] Our research team conducted a non-inferiority online RCT to compare the effect of the crowdsourced video to a health marketing video to promote HIV testing.[24]

In this online RCT, the participants were recruited from popular MSM websites in three regions of China: northern (www.danlan.com), southern (www.yztz.net), and eastern (www.jstz.org). Participants entered the survey by clicking on a banner advertisement or an announcement sent to registered users, which directed them to eligibility screening and consent. Inclusion criteria were:1) being born biologically male; 2) having had anal sex with a man at least once; 3) 16 years or older; 4) never tested for HIV; 5) able to provide a cell phone number for follow up. In total, 721 participants were recruited from 31 provinces in 217 cities in China. Of those 721 individuals, 352 were randomly assigned to the crowdsourced intervention arm and 369 to the health marketing intervention arm. The study protocol was approved by the participating institutional review boards and the study was registered with ClinicalTrials.gov (identifier NCT02248558) prior to trial enrollment.

A text message was sent to participants three weeks after survey completion, asking about HIV test uptake and test results after the intervention. An identical second follow-up text message was sent to non-responders in the fourth week. Of the 721 participants, 624 (87%) replied to the text message. Response (i.e., retention) rates were similar between the crowdsourced group (87%) and health marketing group (86%).[24]

The larger population of interest

The larger population of interest was generated by the following steps. First, we used a national, large-scale, cross-sectional survey dataset on MSM in China[25] collected in 2013 (sentinel surveillance data) as a representative sample of the measurable Chinese MSM population. The data were collected by the Chinese Center for Disease Control and Prevention (C-CDC) aiming to determine the burden of HIV among MSM. A multistage mixed method sampling strategy including snowballing, venue- and Internet-based sampling was used to recruit participants. Although the C-CDC study was not a theoretically random sample of MSM, we consider it representative of the population of interest for the following reasons: 1) consistency between the C-CDC data set and extensive HIV systematic review data;[28] 2) large sample size (42680 MSM) recruited; 3) wide geographic reach with sampling from 107 cities within 30 of 31 provinces or similar administrative structures in China. Individuals in the C-CDC data were included in our study if they met the crowdsourcing RCT inclusion criteria (as described above, excluding the cell phone number criteria). Among the C-CDC sample, 20 428 participants met the RCT inclusion criteria and were included for analysis. We could not acquire information on what proportion of these 20

428 MSM from the C-CDC data use the Internet, but several studies have shown that, in China, 79.1% of MSM seek male sexual partners by Internet,[29] and 93.6% of MSM own a smart phone.[30]

Second, to implement the IPSW approach, we estimated that the larger MSM population of interest was in the range of 980 000 to 4 880 000 individuals. This range was determined by multiplying the following estimates: (i) the estimated whole Chinese MSM population size was 2 000 000 to 10 000 000[31] in 2009, which is the latest available estimated MSM population size in China; (ii) the percentage of Chinese MSM of more than 16 years of age and without HIV testing history in 2009 was 48.8% (95% confidence interval (CI): 36.6 to 61.0%).[32] Therefore, we up-weighted the included C-CDC observations by 980 000 / 20 428 and 4 880 000 / 20 428 to get a low and high approximation of the size of the larger population of interest.

Statistical Methods

IPSW—We start with a description of notation to specify the IPSW estimator. Let N denote the size of the population of interest. Let $S_i = 1$ denote selection of the ith individual from the population of interest into the RCT, and let $S_i = 0$ otherwise. For $S_i = 1$, let $X_i = 1$ denote random assignment to the crowdsourcing arm and $X_i = 0$ to the health marketing arm. Let $\mathbf{Z_i}$ denote a vector of k covariates for individual i. The IPSW was defined as[17]

$$\widehat{W}_i = \begin{cases} \dfrac{\widehat{P}(S_i = 1)}{\widehat{P}(S_i = 1 \mid \mathbf{Z}_i)} & \text{if } S_i = 1 \\ 0 & \text{if } S_i = 0 \end{cases}$$

where $\hat{P}(S_i = 1|\mathbf{Z_i})$ is an estimate of the probability of being selected into the RCT conditional upon the measured covariates $\mathbf{Z_i}$. The estimated marginal probability of being selected into the RCT was used in the numerator of the IPSW in order to ensure that the weighted sample remains the same size as the original sample and to check that the mean of $\hat{W}_i$ is near 1.0. The weights $\hat{W}_i$ are inversely proportional to an estimate of the conditional probability of being selected in the trial. A logistic regression model was used to obtain $\hat{P}(S_i = 1|\mathbf{Z_i})$with main effects and all 2-way interactions of covariates $\mathbf{Z_i}$ included in the model.

After estimating the weights, the RCT data were analyzed with an IPSW linear probability model $Y_i = \beta_0 + \beta_1 X_i$ where Y_i denotes the outcome of interest, defined as Y_i=1 if the ith individual reported getting HIV tested within four weeks following the video intervention, and Y_i=0 otherwise. The estimate of the regression coefficient β_1 was interpreted as an estimate of the difference in the probability of getting HIV tested (crowdsourced – health marketing) in the population of interest.[33] A robust variance estimator was used to construct a Wald confidence interval (CI).[34]

G-formula—The G-formula[20,21,35] for generalizability[36] accounts for non-random RCT participant selection (S_i) by integrating over the covariate distribution $\mathbf{Z_i}$ from the population of interest. Specifically, the G-formula estimator of the difference in the probability of getting HIV tested in the population of interest is

$$\int (\hat{E}[Y_i|X_i = 1, \mathbf{Z}_i = \mathbf{z}, S_i = 1] - \hat{E}[Y_i|X_i = 0, \mathbf{Z}_i = \mathbf{z}, S_i = 1])d\hat{F}(\mathbf{z})$$

where the estimated conditional expectations $\hat{E}[Y_i]X_i = x$, $\mathbf{Z_i} = \mathbf{z}$, $S_i = 1]$ for x = 0,1 are based on data from the RCT and $d\hat{F}(\mathbf{z})$ is the empirical distribution of $\mathbf{Z}$ in the C-CDC data. A general linear model was used to estimate the mean HIV testing outcome conditional on random intervention assignment X_i and covariates $\mathbf{Z}_i$ among the RCT participants. Main effects and all 2-way interactions between covariates $\mathbf{Z}_i$ and the intervention assignment X_i were included in the model. A Wald CI was computed using a bootstrap standard error estimate obtained by resampling with replacement from both the RCT sample and C-CDC sample separately (1000 resamples).

Assumptions[17,18,36]—The IPSW and generalizability G-formula methods rely on the following assumptions: (i) no measurement error; (ii) conditional on measured covariates, the participants selected to the RCT are exchangeable with those who were not selected (i.e., conditional exchangeability); (iii) non-zero probability of being selected into the RCT sample in each strata of covariates; (iv) no interference in both the study population and the larger population of interest; (v) correct model specification. If conditional exchangeability is met, effect measure modification observed in the RCT sample is representative of that in the larger population of interest

Covariate selection—The following three criteria were used to select covariates $\mathbf{Z_i}$ for adjustment[17]: 1) the covariates were associated with RCT participation and were potential modifiers of the effect of the intervention; 2) the covariates were available in both the RCT data and the target sample C-CDC data, and 3) the covariates were deemed of scientific relevance for potential to impact the HIV testing outcome. The variables age, marital status, education, and condom use during last anal sex in the past six months met these criteria. However, the condom use variable was missing for 13.2% and 36.8% of MSM in the RCT and C-CDC data, respectively. Due to this substantial missingness, we only included age, marital status and education in the main analyses. Results including the variable condom use are provided in the supplemental tables. All available covariates were measured by C-CDC using categories (e.g., age bins) and were fit as categories in the analyses.

All analyses were conducted in SAS version 9.2 (Cary, NC) and code is provided in the online supplement (See Supplementary SAS Code).

RESULTS

Compared to the larger population of interest, online RCT participants were more likely to be age 20 or under (32.5% versus 9.8%), well educated (70.5% versus 40.8%), and never married (85.3% versus 67.7%). After weighting the crowdsourced RCT using IPSW, the distribution of covariates of the weighted RCT was close to the larger population of interest (Table 1 and S1).

In our analysis, first we replicated the main results of the crowdsourced RCT.[24] The estimated difference in proportions of self-reported HIV testing within 4 weeks after

watching the assigned video between two arms (crowdsourced – health marketing) was 2.1% (Wald 95% CI, −5.4% to 9.7%). The hypothesis that the crowdsourced video was not inferior to the health marketing video to promote HIV testing was not demonstrated; the CI included values below the non-inferiority margin of −3% (Table 2).

Additionally, we assessed effect measure modification of the difference in proportions tested between the two arms (Table 2 and S2). There was variability in the subgroup effect estimates, particularly for the youngest and oldest groups and by marital status (Figure 1). However, these analyses of marginal effect measure modification (one covariate at a time) do not assess potential higher-order, joint effect measure modification.

In the IPSW analysis (including the covariates in Table 1 and all their 2-way interactions), non-inferiority was also not demonstrated. The estimated difference in proportions was on the opposite side of the zero null effect as compared to the original RCT result (difference in proportions: −2.6%; 95% CI: −14.2 to 8.9%) (Table 3). As anticipated, the IPSW robust CI was wider than the RCT original CI (CI widths of 23.1% and 15.1%, respectively). IPSW results were almost the same under the two approximations of population size. The mean of weights W_i was 0.996 (SD=1.4). In the G-formula analysis, the estimated difference in proportions was 2.7 (95% CI: −10.7 to 16.2), again showing that non-inferiority was not demonstrated (Table 3). The G-formula bootstrap CI was slightly wider (width 26.9%) than the IPSW robust CI.

In the supplement, we provide IPSW and G-formula results with the condomless sex variable included in the covariate set $\mathbf{Z}_i$. Missing data were excluded, thus we make a strong assumption that report of condomless sex was missing completely at random. The IPSW estimated effect was nearly unchanged by inclusion of the condomless sex variable (Table S3) but the standard error was considerably larger due to loss of data (i.e., smaller evaluable RCT sample size). The G-formula result appeared somewhat sensitive to inclusion of the condomless sex variable. In the supplemental analysis that incorporates the condomless sex variable, the IPSW and G-formula results were in close agreement with one another and were both on the opposite size of the zero null effect as compared to the original RCT result. Non-inferiority was not demonstrated in any of our analyses.

DISCUSSION

In our study, we applied IPSW and G-formula methods to generalize inferences from a crowdsourced online RCT to a specified population of interest. Both methods supported the original online RCT conclusion when the crowdsourced RCT was quantitatively generalized to a specified larger population of interest. These two different approaches have different model assumptions, so we have a greater degree of confidence in the results when the two estimates are similar. This study adds to the current literature by using two different model-based methods to examine the generalizability of an online, crowdsourced RCT.

Our study suggests that properly planned and conducted online RCT results can be generalizable to a specific larger population of interest. Both IPSW and G-formula generalizability methods showed that non-inferiority was not demonstrated; conditional on

several key analysis assumptions, these results may be generalized to the larger China MSM population of interest. The estimated difference from the G-formula analysis was quite close to the original RCT result. The estimated difference from IPSW analysis was not drastically different from the original RCT result, but was on the opposite side of the zero null effect. The disagreement between the G-formula and the IPSW could be due to one or both of the assumed parametric models being incorrectly specified.

For both the IPSW and G-formula estimates, the estimated standard error was larger than in the original RCT analysis, indicating a loss in precision when the RCT results were generalized to a larger population. In general, loss of precision is expected with increasing differences between participant characteristics in the RCT sample compared to the population of interest.[17,18]

Our study has several limitations. First, for the assumption of no measurement error[17,37], we acknowledge that self-reported data, like data for HIV testing and condomless anal sex, are susceptible to reporting bias. Secondly, we only had a small number of measured covariates that could be compared between the online RCT survey and CDC survey data sets, and thus conditional exchangeability may be violated. Other possible effect measure modifiers, which mean potential for unmeasured covariates meeting the covariate selection criteria (as described above, excluding the second criteria), may have been missed such as annual income and disclosure of sexual orientation. However, the C-CDC data was thus far the most comprehensive, largest and only national data available to us. Thirdly, we could not acquire information on Internet access for the C-CDC data. Nonetheless, the IPSW and G-formula generalizability results do adjust for age, which may potentially mitigate bias due to Internet access not being included among the observed covariates since different age groups are known to have different Internet usage rates.[29] Collecting information on Internet access from the population of interest is recommended in general to facilitate generalizability of online RCTs. Lastly, the variable of condom use during last anal sex in last six months was excluded in our primary analysis due to substantial missing data. Sensitive behavioral data such as this can be difficult to measure.

In conclusion, we used a crowdsourced online RCT and C-CDC data as an example to examine the generalizability of an online RCT with two different methods. We demonstrate the feasibility of a generalizability analysis in this context and provide example SAS code. Considering the advantage of efficiently recruiting a large, high-risk population, our results support that a properly planned and conducted online RCT is an effective design for estimating the effects of intervention or treatment. Examining the generalizability of online RCTs is an important step before the interventions or treatments are scaled up to the population of interest. Future generalizability analyses can be facilitated by harmonizing covariate data collected in RCTs and in large samples of the population of interest (e.g., surveillance or cohort studies).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all the staff members at SESH Global, Yan Cui, Wei Guo, Qianqian Qin and other members at the Chinese Center for Disease Control and Prevention who contributed. We thank Catherine Lesko, Stephen Cole, Jess Edwards, and Camden Bay for their advice.

Funding

This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID), US National Institutes of Health(1R01AI114310); University of North Carolina(UNC)–South China STD Research Training Centre (Fogarty International Center grant number 1D43TW009532); UNC Center for AIDS Research(NIAID grant number 5P30AI050410); This publication was supported by the National Center for Advancing Translational Sciences at the National Institutes of Health (grant number UL1TR001111).

References

1. Mathieu E, Barratt AL, McGeechan K, Davey HM, Howard K, Houssami N. Helping women make choices about mammography screening: an online randomized trial of a decision aid for 40-year-old women. Patient education and counseling. 2010; 81:63–72. [PubMed: 20149953]
2. Naslund JA, Aschbrenner KA, Marsch LA, McHugo GJ, Bartels SJ. Crowdsourcing for conducting randomized trials of internet delivered interventions in people with serious mental illness: A systematic review. Contemp Clin Trials. 2015; 44:77–88. [PubMed: 26188164]
3. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc. 2013; 20:568–76. [PubMed: 23065196]
4. Pequegnat W, Rosser BS, Bowen AM, et al. Conducting Internet-based HIV/STD prevention survey research: considerations in design and evaluation. AIDS and Behavior. 2007; 11:505–21. [PubMed: 17053853]
5. Shin E, Johnson TP, Rao K. Survey mode effects on data quality: Comparison of web and mail modes in a US national panel survey. Social Science Computer Review. 2012; 30:212–28.
6. Bech, M., Kristensen, MB., editors. Survey Research Methods. Citeseer; 2009. Differential response rates in postal and web-based surveys among older respondents.
7. Denscombe M. Item non-response rates: a comparison of online and paper questionnaires. International Journal of Social Research Methodology. 2009; 12:281–91.
8. Kwak N, Radler B. A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. Journal of official statistics. 2002; 18:257.
9. Blas MM, Alva IE, Carcamo CP, et al. Effect of an online video-based intervention to increase HIV testing in men who have sex with men in Peru. PLoS One. 2010; 5:e10448. [PubMed: 20454667]
10. Côté J, Cossette S, Ramirez-Garcia P, et al. Evaluation of a Web-based tailored intervention (TAVIE en santé) to support people living with HIV in the adoption of health promoting behaviours: an online randomized controlled trial protocol. BMC Public Health. 2015; 15:1042. [PubMed: 26458508]
11. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol. 2004; 55:803–32. [PubMed: 14744235]
12. Göritz AS. Incentives in web studies: Methodological issues and a review. International Journal of Internet Science. 2006; 1:58–70.
13. Hirshfield S, Chiasson MA, Joseph H, et al. An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. PLoS One. 2012; 7:e46252. [PubMed: 23071551]
14. MIETTINEN OS. Standardization of risk ratios. American Journal of Epidemiology. 1972; 96:383–8. [PubMed: 4643670]
15. Rothman, KJ., Greenland, S., Lash, TL. Modern epidemiology. Lippincott Williams & Wilkins; 2008.
16. Buchanan, AL., Hudgens, MG., Cole, SR., et al. Generalizing Evidence from Randomized Trials using Inverse Probability of Sampling Weights. the Berkeley Electronic Press; 2015.

17. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations the actg 320 trial. American journal of epidemiology. 2010; 172:107–15. [PubMed: 20547574]
18. Bengtson AM, Pence BW, Gaynes BN, et al. Improving Depression Among HIV-Infected Adults: Transporting the Effect of a Depression Treatment Intervention to Routine Care. J Acquir Immune Defic Syndr. 2016; 73:482–8. [PubMed: 27668804]
19. Greenhouse JB, Kaizar EE, Kelleher K, Seltman H, Gardner W. Generalizing from clinical trial data: a case study. The risk of suicidality among pediatric antidepressant users. Stat Med. 2008; 27:1801–13. [PubMed: 18381709]
20. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period —application to control of the healthy worker survivor effect. Mathematical Modelling. 1986; 7:1393–512.
21. Pearl, J. Causality: models, reasoning, and inference. Cambridge university press; 2009.
22. Westreich D, Cole SR, Young JG, et al. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. Stat Med. 2012; 31:2000–9. [PubMed: 22495733]
23. MA, HJMR. Causal Inference. Boca Raton: Chapman & Hall/CRC; 2017.
24. Qin Q, Tang W, Ge L, et al. Changing trend of HIV, Syphilis and Hepatitis C among Men Who Have Sex with Men in China. Scientific Reports. 2016; 6
25. Qin Q, Tang W, Ge L, et al. Changing trend of HIV Syphilis and Hepatitis C among Men Who Have Sex with Men in China. Sci Rep. 2016; 6:1–10. [PubMed: 28442746]
26. Zhang Y, Kim JA, Liu F, et al. Creative contributory contests to spur innovation in sexual health: 2 cases and a guide for implementation. Sex Transm Dis. 2015; 42:625–8. [PubMed: 26462186]
27. Howe J. The rise of crowdsourcing. Wired magazine. 2006; 14:1–4.
28. Zhang L, Chow EP, Jing J, et al. HIV prevalence in China: integration of surveillance data and a systematic review. The Lancet infectious diseases. 2013; 13:955–63. [PubMed: 24107261]
29. Liu Y, Wang J, Qian HZ, et al. Seeking Male Sexual Partners via Internet and Traditional Venues among Chinese Men Who Have Sex with Men: Implications for HIV Risk Reduction Interventions. Aids & Behavior. 2016; 20:2222. [PubMed: 27000143]
30. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay Apps for Seeking Sex Partners in China: Implications for MSM Sexual Health. Aids & Behavior. 2015; 19:941–6. [PubMed: 25572834]
31. Tang S, Tang W, Meyers K, Chan P, Chen Z, Tucker JD. HIV epidemiology and responses among men who have sex with men and transgender individuals in China: a scoping review. BMC Infect Dis. 2016; 16:588. [PubMed: 27765021]
32. Chow E, Wilson D, Zhang L. The rate of HIV testing is increasing among men who have sex with men in China. HIV medicine. 2012; 13:255–63. [PubMed: 22252151]
33. Agresti, A., Kateri, M. International encyclopedia of statistical science. Springer; 2011. Categorical data analysis; p. 206-8.
34. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168:656–64. [PubMed: 18682488]
35. Bareinboim, E., Pearl, J. Transportability of casual effects: Completeness results; AAAI Conference on Artificial Intelligence; North America. 2012;
36. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiology. 28(4):553–561.
37. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics. 2011; 67:1406–13. [PubMed: 21627630]

What is already known on this subject?

Investigators increasingly use online methods to recruit participants for randomized controlled trials (RCTs). The extent to which participants recruited online represent populations of interest is unknown. Statistical methods have been developed for generalizing RCT findings to populations of interest. However, to date, these types of methods have not been employed to generalize results of online RCTs.

What does this study add?

- Globally, this is one of very few studies evaluating the generalizability of an online RCT.
- Our study was the first, or one of the first, applications of the G formula to examine the generalizability to RCT data.
- Conducting generalizability analysis of an online RCT is feasible.
- Examining the generalizability of online RCTs is an important step before an intervention is scaled-up to a population of interest.

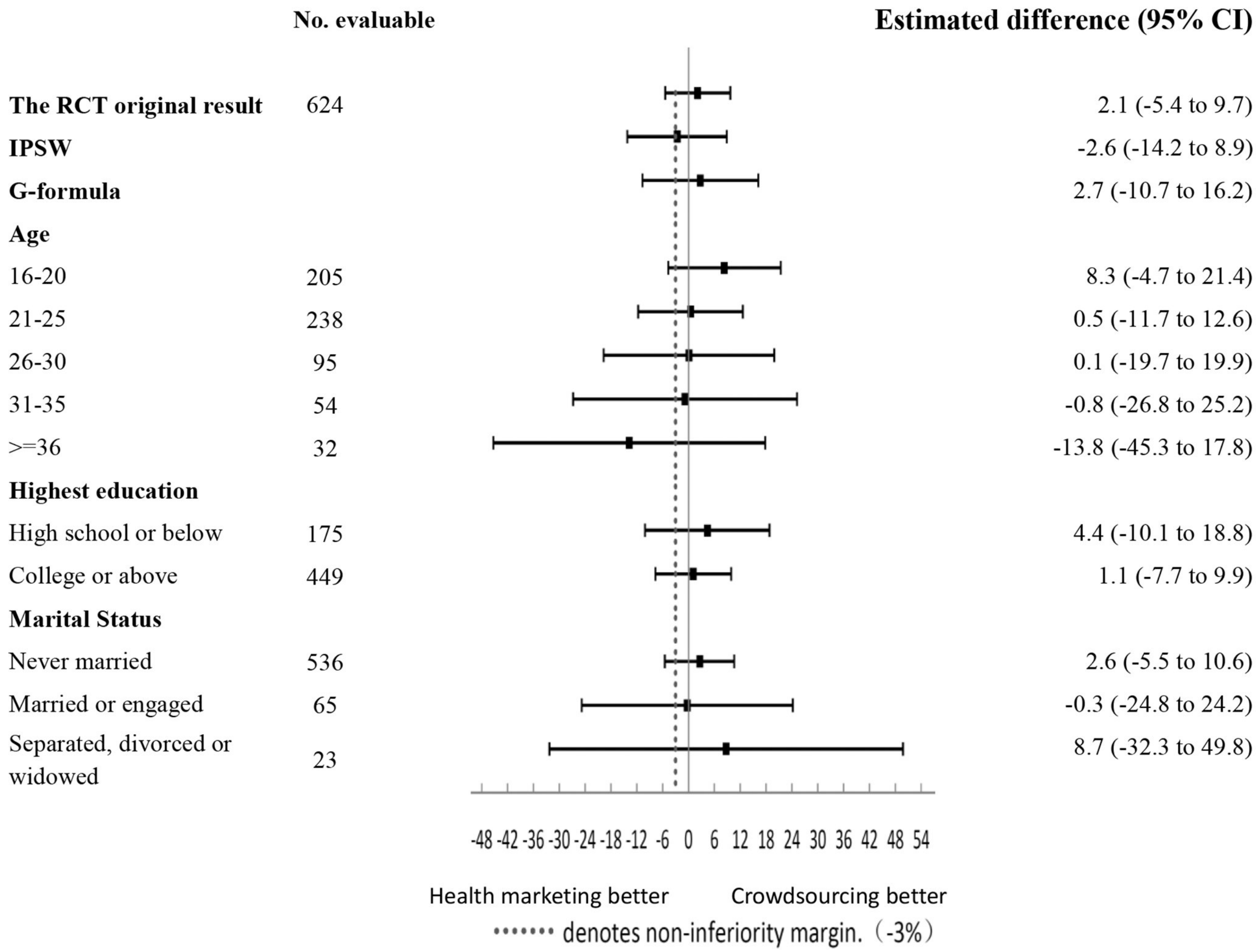

Figure 1.

The results of original RCT, IPSW, G-formula and sub-analyses

Table 1

Characteristics of Covariates in RCT in China in 2014, Up-weighted C-CDC Study in China in 2013 and Inverse Probability of Sampling Weighted RCT Population[a]

	RCT (n=721)		Up-weighted C-CDC (N_{total}=980000)		weighted RCT	
	Freq	**%**	**Freq**	**%**	**Freq**	**%**
Age						
16–20	234	32.5	96135.8	9.8	70.9	9.9
21–25	266	36.9	293966.5	30.0	214.3	29.8
26–30	112	15.5	232048.5	23.7	174.1	24.2
31–35	64	8.9	141999.1	14.5	104.6	14.6
>=36	45	6.2	215850.2	22.0	154.6	21.5
missing	-	-	-	-	-	-
Highest education						
High school or below	213	29.5	579594.1	59.2	423.8	59.0
College or above	508	70.5	399830.8	40.8	294.6	41.0
missing	-	-	*575.1*	*0.06*	-	-
Marital Status						
never married	615	85.3	662694.5	67.7	487.2	67.8
married/engaged	81	11.2	256681.5	26.2	190.3	26.5
separated/divorced/widowed	25	3.5	60192.7	6.1	40.9	5.7
missing	-	-	*343.3*	*0.04*	-	-

[a]Percentages were calculated among non-missing observations, and percent missing among total n is also shown in italics.

Table2

The Original RCT Complete-case HIV Testing Results and Sub-analyses in China in 2014 (n=624)

Subgroup	N	Tested/n(%)		Estimated Difference(%)	S.E.(%)	95%CI(%)	P value
		crowdsourced	health marketing				
overall	624	114/307(37.1)	111/317(35.0)	2.1	3.8	−5.4 to 9.7	0.58
age							0.75
16–20	205	42/106(39.6)	31/99(31.3)	8.3	6.7	−4.7 to 21.4	
21–25	238	43/121(35.5)	41/117(35.0)	0.5	6.2	−11.7 to 12.6	
26–30	95	16/41(39.0)	21/54(38.9)	0.1	10.1	−19.7 to 19.9	
31–35	54	10/26(38.5)	11/28(39.3)	−0.8	13.3	−26.8 to 25.2	
>=36	32	3/13(23.1)	7/19(36.8)	−13.8	16.1	−45.3 to 17.8	
Highest education							0.71
High school or below	175	37/89(41.6)	32/86(37.2)	4.4	7.4	−10.1 to 18.8	
college or above	449	77/218(35.3)	79/231(34.2)	1.1	4.5	−7.7 to 9.9	
Marital Status							0.93
never married	536	98/271(36.2)	89/265(33.6)	2.6	4.1	−5.5 to 10.6	
married or engaged	65	12/27(44.4)	17/38(44.7)	−0.3	12.5	−24.8 to 24.2	
separated, divorced or widowed	23	4/9(44.4)	5/14(35.7)	8.7	20.9	−32.3 to 49.8	

Table 3

The Generalized Effect of Crowdsourced Versus Health Marketing Video Upon HIV Testing[a]

Method	Difference in proportions,%	Standard Error, %	95% CI,%
IPSW[b]	−2.6	5.9	−14.2 to 8.9
G-formula	2.7	6.9	−10.7 to 16.2

[a]The covariates included in this analysis were age category, health education, and marital status.

[b]IPSW=inverse probability of sampling weight