

TODD C. HELMUS, KURT KLEIN

Assessing Outcomes of Online Campaigns Countering Violent Extremism

A Case Study of the Redirect Method

KEY FINDINGS

- Virtually all of the reviewed evaluations of campaigns aimed at countering violent extremism (CVE) focused on analysis of reach, views, and engagement, which do not gauge impact on audience attitudes or behavior.
- “Closed” evaluation designs enable researchers to assess changes in audience attitudes through the use of participant surveys and random assignment of CVE content.
- Other approaches to evaluation include (1) measuring audience responses to “calls to action” social media campaigns that encourage audience members to engage in an observable behavior, such as calling specific phone lines or clicking on particular links; (2) assessing audience comments; and (3) conducting A-B testing, in which media planners compare audience reactions to two or more versions of online content.
- After examining data for a specific online CVE campaign called the Redirect Method—which used Google’s AdWord technology to serve CVE content to users searching for violent jihadist and violent far-right content—AdWord analytics data suggested that the campaign was most successful in placing Google ads and CVE video content in front of users who searched for violent jihadist content.
- To move beyond process evaluations and assess real impact, a more data-driven and experimental approach is needed.

While the number of programs dedicated to countering violent extremism (CVE) has grown in recent years, a fundamental gap remains in the understanding of the effectiveness of such programs. A 2017 RAND Corporation report documented that only a handful of such programs have been subject to rigorous evaluations of impact (Helmus et al., 2017). Such evaluations are critical because they help ensure that programming funds are dedicated to the most-effective efforts. Evaluations also play a critical role in helping individual programs improve the quality of service provision.

CVE campaigns are increasingly moving to the online space, and such campaigns are no less in need of assessment and evaluation (Reynolds and Tuck, 2016). However, assessment of online CVE campaigns presents some unique opportunities and challenges. Many such

campaigns—whether implemented on YouTube, Facebook, or Twitter—have at their disposal a tailored set of analytics provided by the social media platform being employed. Such analytics can identify the number of individuals that view, like, share, or comment on campaign content. Some platforms even provide information about campaign audience demographics—a wealth of data that can greatly aid campaign implementation.

However, the unique nature of social media can impede conventional approaches to assessing program impact. For example, exposure to social media content is highly self-selective (users are not passively exposed to material; they choose which content to consume), and this can easily bias evaluations. In addition, the rapid flow of social media information across browser screens can reduce the accuracy of exposure recall (Andersen, de Vreese, and Albaek, 2016; Niederdeppe, 2016; Niederdeppe, 2014; de Vreese and Neijens, 2016).¹

In this brief report, we investigate some of these and other evaluation challenges for online campaigns. We seek to help a broad array of CVE campaign planners and evaluators better assess their social media-based CVE campaigns. To do this, the report first describes prior evaluations of online CVE campaigns. As expected, our review indicated that few of these evaluations actually examine campaign impact on target audience attitudes or behaviors and focus instead on reach and audience engagement. To address this limitation, we then reviewed different methods for evaluating the impact of online campaigns on attitudes and behavior, which we drew from public health literature in which researchers conducted evaluations for online campaigns seeking to promote smoking cessation, reduce youth drinking, and achieve other such outcomes. We specifically provide an informative

review of the different methodological approaches for testing impact of online campaigns with the hopes that readers can draw on these approaches to help design their own online CVE evaluations.

We then take an in-depth look at a specific online CVE campaign called the Redirect Method (Redirect Method, undated), which was developed by the “think-do tank” Jigsaw (a subsidiary of Google’s parent company Alphabet Inc.) in collaboration with the Gen Next Foundation, Moonshot CVE, and others. Taking advantage of the technology behind Google AdWords, this method has been used to identify potential Islamic State of Iraq and al-Sham (ISIS) recruits through their Google searches and exposed them to ads linking to curated YouTube videos debunking ISIS recruiting themes (Redirect Method, undated). We examine select AdWord analytics data from a Redirect Method campaign targeting both violent jihadist and violent far-right extremists in the United States. Then, drawing on the previously reviewed approaches for assessing the impact of online campaigns, we provide recommendations for how researchers can conduct more-rigorous evaluations of an intervention’s impact on target audience attitudes and behavior.

Review of Online CVE Interventions

We conducted a literature search on Google Scholar to generate a list of promising bibliographies. In reviewing the bibliographies, we sought to identify evaluations of online campaigns with a CVE focus and we distilled our list to seven studies. (The appendix provides a summary of the methodology and findings of the studies identified for this review.)

These studies examined a wide array of social media or web-based programs. Louis Reynolds and Henry Tuck at the Institute for Strategic Dialogue (ISD) reviewed a series of online campaigns that might be considered prototypical influence efforts. The “Average Mohamed” campaign, for example, disseminated five animated videos that drew on Islamic principles pertaining to peace to “counter the ideology of Islamist extremist groups” while the ExitUSA campaign targeted far-right extremists

Abbreviations

CTR	click-through rate
CVE	countering violent extremism
ISD	Institute for Strategic Dialogue
ISIS	Islamic State of Iraq and al-Sham

with four videos highlighting the experience of former extremists (Reynolds and Tuck, 2016). Two studies focused on an online intervention that had former extremists reach out directly to extremist users via Facebook Messenger (Frenett and Dow, 2014; Davey, Birdwell, and Skellett, 2018) while others examined counterextremist speech on Facebook (Bartlett and Krasodomski-Jones, 2015). Only one of these studies was published in a peer-reviewed journal (Al-Rawi, 2013).

By and large, the studies we reviewed examined basic process characteristics of the communication campaign. Academic experts generally partition communication campaign research into three categories: formative, process, and summative or impact evaluations (Chung, 2016). Formative studies, conducted in the development phase of the campaign, assess strengths and weaknesses of the campaign materials. Process evaluations are conducted during the campaign and assess degrees of reach, distribution, audience reception, and engagement. This is the sweet spot addressed by social media analytics.² Finally, summative or impact evaluations examine changes in audience behavior, attitudes, and perceptions.

Virtually all of the reviewed CVE evaluations focused on analysis of reach, views, and engagement (likes, shares, comments) and thus could best be defined as process evaluations. ISD, for instance, conducted four brief case studies of CVE social media assessments (Silverman et al., 2016; Reynolds and Tuck, 2016). One such case study examined the “Extreme Dialogue” campaign that disseminated short, emotive films on YouTube, Facebook, and Twitter to dispel extremist myths and encourage “empathy with and understanding of the ‘other.’” The Facebook campaign reached the feeds of 362,500 unique users, 1,835 of whom clicked on the links (a 0.5-percent click-through rate [CTR]). According to the authors, the low CTR suggested that the ads might have either reached the wrong audience—those not interested in clicking the content—or relied on unappealing advertisements.³ But half of those who did watch the video content responded by liking, sharing, or commenting on the campaign, which suggested that those who watched the campaigns were sufficiently engaged by them. Indeed, a limited analysis of the demographics of

those who engaged with the campaign suggested that the YouTube campaign did a better job of reaching the target audience of those between the ages of 18 and 24 than did the Facebook campaign, whose engaged audience was older. However, little is known about the campaign’s impact on audience attitudes or behavior.

Several campaigns analyzed comments. Al-Rawi, for example, analyzed 281 comments from a YouTube-based counterterrorism campaign and determined that many commenters expressed suspicion toward a campaign presumably funded by a western government. Commenters also suggested that the campaign was preachy in tone and appeared to dictate basic facts about Islam to the public (Al-Rawi, 2013). Such a process measure, if used early in the campaign, can presumably help identify needed improvements to video content.

In contrast, Silverman et al. (2016) consider at least some comments as a form of impact measure. They specifically highlight the cases of eight individuals who reached out to the Life After Hate group for information and support following exposure to an ExitUSA campaign that targeted far-right extremists with videos depicting the storylines of former violent extremists. One individual, for example, identified as a white supremacist who “still gets old feelings and thoughts.” As the authors note, “These messages are perhaps the most direct evidence possible of counter narrative campaigns having impact” (Silverman et al., 2016, p. 39) They attribute this success partly to the ExitUSA team’s diligence in consistently responding to comments and partly to the team’s background as former extremists. (We will discuss the value of comments as an impact measure more broadly in a later section.)

Frenett and Dow (2014) and Davey, Birdwell, and Skellett (2018) analyzed dialogue between extremists on Facebook and the former extremists who reached out via Facebook Messenger. Davey, Birdwell, and Skellett used analyses of such conversations to assess not only engagement and response but also what they call “potential positive impact,” which they exemplified as an online extremist expressing interest in taking the conversation offline, indicating that the conversation might have changed attitudes or had a positive effect on negative online behaviors. The

ability to cultivate a sustained dialogue provides additional opportunities to assess effects that are not otherwise visible in the comments posted on social media channels.

In summary, evaluation of online CVE campaigns is a new and burgeoning phenomenon as CVE implementers increasingly use social media in this way. The field is still immature; very few systematic evaluations are published, most of which are brief, case-study analyses—and most of those rely on analytics provided by the platform, which typically address such process measures as reach, likes, and engagement but do not assess campaign-induced changes in audience attitudes or behavior. As a result, little is known about the broader impact of the campaigns. One campaign did use comments as a measure of impact; the merits of this approach are addressed later in this report.

Assessing Impact and Audience Reaction to Online Content

Online CVE campaigns have rarely assessed impact or audience perceptions of content. In this section, we identify different methodological approaches for impact evaluations of online campaigns. Our analysis draws on a brief review of studies assessing the impact of social media-based public health campaigns that seek to promote, for instance, smoking cessation, condom usage, testing for sexually transmitted diseases, and reductions in youth drinking. We also supplemented this review with interviews with select experts.⁴

Open Versus Closed Designs

A recently published systematic review on social media evaluation methodologies points to the

Online CVE campaigns have rarely assessed impact or audience perceptions of content.

differential use of open versus closed evaluation designs (Lim et al., 2016). Open designs are essentially evaluations conducted on live social media campaigns. These designs can draw on all the process measures available from the platform analytics, such as impressions, reach, and engagement (likes, shares, comments). Impact evaluations for these campaigns typically consist of pre-post surveys, although the self-selective nature of social media use (i.e., users choose which content to consume) can easily bias comparisons between audiences that are exposed to a campaign and audiences that are not.⁵

In contrast, closed designs draw on a specially recruited population of participants that are often (though not necessarily) randomly assigned to either a treatment or control condition. Researchers then formally expose the participants in the treatment condition to the social media content and conduct follow-on surveys or other tests to assess whether exposure to the social media content produced its intended effect.⁶ Participants can be recruited via online methods (e.g., Facebook surveys, Amazon Mechanical Turk) or offline ones (e.g., via newspaper advertisements, advertisements on college campuses, etc.).

These approaches offer different strengths and weaknesses. Closed designs allow more-rigorous evaluation design for testing the impact of online content because participants can be randomly assigned to treatment exposure or control conditions and surveys can be administered more systematically to test treatment effects over time.⁷ However, closed designs lack ecological validity—participants do not necessarily consume the provided content in the same natural way that they would by scrolling through their social media feeds. In contrast, evaluators of open campaigns can use a platform's analytics to test reach and reactivity to the campaign and to evaluate other process-oriented metrics.

Whether to use an open or closed campaign depends on several factors, such as intervention type, target population, and available resources. For example, a campaign that targets extremist populations could find it difficult (though not necessarily impossible) to recruit participants into closed panel studies.⁸ We believe that highly resourced campaigns would benefit from employing both designs.

Researchers could use closed designs to test whether the program's content produces intended audience reactions (e.g., whether participants like the ads) and to evaluate short- and long-term changes in attitudes and behavior. Campaign planners could then conduct an open campaign to deliver content to at-risk audiences, and evaluators could review platform analytics to assess reach and reactivity.

Surveys

Surveys are one of several options to consider for evaluating the impact of CVE social media campaigns. Major, population-based opinion poll surveys are expensive and might fail to adequately identify populations exposed to the campaign's online content, but other options exist that are more cost-effective. Intercept surveys are one such option. With this method, researchers seek to question respondents at a particular location. For example, a study targeting a local lesbian, gay, bisexual and transgender (LGBT) community with an online anti-smoking campaign conducted surveys in local LGBT bars (Fallin et al., 2015).⁹ A CVE campaign seeking to reach and influence community high-schoolers could conduct pre-post surveys in a particular school. It might also be possible to introduce a survey response option following exposure to YouTube videos.

Studies using closed evaluation designs can recruit panel participants via various web and social media channels, such as Facebook, Google AdWords, and Amazon Mechanical Turk (Antoun et al., 2016).¹⁰ Historically, the obvious benefit of using Google AdWords or Facebook is the ability to specially target hard-to-reach populations, such as those with extremist views.¹¹

A major value of closed evaluation panel surveys is that they can estimate both impact and reaction to content. Mowery et al. (2016), for example, recruited a research panel to assess anti-smoking videos created by the Centers for Disease Control and Prevention. They asked participants to rate the video content on a host of variables. Had they conducted follow-up surveys at multiple times, they could have tested the long-term impact of the videos. However, their one-time survey

asked participants about their intention to quit and about their broader attitudes regarding quitting. The survey also featured several questions about audience reaction, such as whether the ads grabbed attention, gave reasons to quit tobacco, were memorable, and were engaging. In this sense, the study had components that achieved formative, process, and impact objectives. The authors concluded that “[o]nline message testing is an efficient way to test ads, improve program effectiveness, and protect investments in public health marketing campaigns” (Mowery et al., 2016, p. 179).

One critique of using surveys is that it can be difficult to construct survey questions that effectively measure phenomena that are as complex and sensitive as radicalization. Responses to questions about extremist views and attitudes can suffer obvious social desirability bias. Participants also might worry about how their data will be used and whether their responses will be kept confidential. The Asia Foundation provides a helpful primer about conducting surveys in support of CVE programs (Nanes and Lau, 2018). They identify several helpful steps designed to improve responses to sensitive questions, such as providing detailed descriptions of how data will be used, policies for confidentiality, and recommendations for moderating or softening questions and depersonalizing questions by asking about viewpoints of “your neighbor.” (An example would be, “to what extent would you say that your community sympathizes with the Klu Klux Klan?”)

Nanes and Lau (2018) also recommend the use of randomization techniques, such as the following:

- **Forced-choice questions** enforce randomization by asking participants to flip a coin and, if heads, answer truthfully; if tails, answer yes regardless of the truth.
- **Endorsement experiments** ask participants whether they support a particular and relatively inane policy and some participants are incidentally informed that an extremist group supports the policy.¹²
- **List experiments**, according to the Asia Foundation report, seek to protect the privacy of a respondent by “including the sensitive

item of interest in a list of innocuous items and then asking the respondent to report only how many of those items they endorse” (Nanes and Lau, 2018, p. 43).¹³

These experiments afford respondents a level of anonymity in answering high-risk questions and require a statistician to analyze data and make inferences of support for the sensitive item being assessed.

Call to Action

Some communication campaigns incorporate a “call to action” behavior into their messaging that can then be easily measured as an indicator of success. One anti-smoking campaign, for example, called on audience members to contact a smoking quitline. The campaign then used the tabulation of those calls as a measure of success (Clayforth et al., 2014). CVE campaigns could use “calls to action” in several ways, such as asking participants to call a mental health line, share posts with their followers, contribute to a Twitter hashtag, click through to an external website, or share stories of walking away from extremism. The key is that the desired action should be stipulated in advance of the campaign and the request should be clearly communicated in the campaign’s content.

Comments

Many online platforms provide opportunities for audience members to comment on the campaign’s content. Facebook posts and videos provide a link for audience members to make comments. YouTube videos also provide space for audience comments. The obvious benefit of comments is that they can serve as a window into the perceptions of a campaign’s audience. The primary way that comment behavior is parsed is to analyze comments posted directly to a campaign’s social media site (e.g., YouTube or Facebook). Such comments can be analyzed in several different ways. For instance, evaluators can hand-code the comment data, as was done with the ExitUSA campaign evaluation, when eight commenters reached out to Life After Hate asking for information and support. An alternative approach is to conduct stance analysis, in which an investigator analyzes comment text to determine whether the author of a text is supportive, opposed, or neutral

regarding a particular topic. Machine learning tools can help with such efforts because customized computer algorithms, often with the help of human supervision, can learn from and make predictions on data-coding categories and can subsequently code sentiment at scale (Sobhani, Mohammad, and Kiritchenko, 2016). One study examined 12,161 comments on 395 YouTube videos that were either pro-anorexia or opposed to pro-anorexia messaging. The authors found that commenters for the anti-pro-anorexia videos exhibited greater degrees of positive sentiment in their comments (Oksanen et al., 2015).

However, analysis of user comments is not infallible. First, comments typically represent only a very small minority of users. The eight individuals who reached out to ExitUSA planners should be placed in the context of the 212,051 individuals exposed to the campaign on Facebook. Second, commenting is a self-selective behavior. Evaluators can easily identify the eight individuals who are seeking help, but these data shed little light on the vast majority of non-commenting audience members. Comments that attract the attention of analysts also might not represent the campaign’s target audience. A campaign seeking to target at-risk youth might attract attention from mainstream society whose comments praising the campaign are essentially meaningless. Thus, evaluators need to exercise great caution in using comments to draw broad conclusions about a campaign’s overall impact, although they can certainly use comments to identify hypotheses about outcomes, identify anecdotes of success, and inform campaign course corrections.

A-B Testing

A-B testing is increasingly seen as a placeholder for impact evaluations. In a marketing context, A-B testing involves media planners running two or more versions of advertisements simultaneously and comparing CTRs to determine which version generates the most clicks. This is an effective way to test advertisement copy because the goal is to elicit a click. However, A-B testing is also used to test the merits of video or other promotional content that the ads link to. It is presumed that videos eliciting longer watch times are more effective. At least one social media

marketing expert said that some firms are seeking to minimize costs and using A-B testing in lieu of costlier focus groups or usability testing.¹⁴ However, it would likely be a mistake to assume that A-B testing can effectively identify the most influential content. For example, a video that won out on multiple A-B tests and elicited long watch times might not effectively influence an audience to adopt the attitudes or behavior desired by the campaign.

The Redirect Method

The Redirect Method has received significant attention in the press as being a potentially effective approach to decreasing the availability of online content that increases the appeal of violent extremist groups, promotes radicalization, and might make viewers more susceptible to recruitment (Greenberg, 2016). As previously noted, Redirect's approach uses Google AdWord technologies to identify individuals searching for violent extremist content on Google and exposes those individuals to an advertisement in their search results that links to counternarrative videos.

The Redirect Method was originally piloted between August 2015 and March 2016. The project identified videos that sought to discredit ISIS narratives with potential sympathizers. The most-engaging videos were organized and showcased in specially designed playlists curated for the campaign. These playlists were then directed globally at individuals whose search patterns suggested positive sentiment toward ISIS. Over the course of eight weeks, these campaigns reached a combined estimate of 320,906 unique individuals by exposing them to a total of 500,070 minutes of watched content (Redirect Method, undated).¹⁵

Gen Next Foundation sought to apply this method to groups in the United States, with the idea of using the approach to target not only violent jihadists but also far-right violent extremists (specifically Neo-Nazi and violent white supremacist movements). CVE practitioners and government funders had previously devoted little attention to the violent far right despite the disproportionately high number of attacks committed by far-right adherents.¹⁶ In this sense,

Redirect's approach uses Google AdWord technologies to identify individuals searching for violent extremist content on Google and exposes those individuals to an advertisement in their search results that links to counternarrative videos.

Gen Next's focus was farsighted, especially in light of events at the Unite the Right March in Charlottesville and the attack at the Tree of Life Synagogue in Pittsburgh (Robertson, Mele, and Tavernise, 2018).

The Redirect Method campaign was funded by the Gen Next Foundation and implemented by Moonshot CVE. The campaign consisted of two separate arms, one focused on targeting violent jihadist extremists and the other targeting violent far-right extremists. The primary campaign that targeted all 50 states was conducted for a four-month period for the violent jihadist campaign and for a three-month period for violent far-right campaign.¹⁷ Moonshot then identified ten state-level counties whose users generate the highest number of extremist searches per capita for each of the two extremist groups and focused a subsequent monthlong campaign on each of those states.

According to interviews with representatives of both Moonshot and Gen Next Foundation, the primary goal of the intervention was to "prevent unobstructed access to extremist content." This was defined specifically as getting advertisements

in the results of violent extremist Google searches. Obviously, ensuring an ad appears where it is supposed to appear is a prerequisite to any subsequent engagement with linked video content. However, another belief was expressed: that the ads themselves help “pierce” a radicalization journey that is often “uninterrupted.” Moonshot CVE specifically defined this goal as a search impression result of 75 percent for both the violent jihadist campaign and the violent far-right campaign. Another goal was to serve counterradicalization video content to viewers.

We analyzed the results for the campaign implemented in all 50 states. To match the intended goals of the campaign, this analysis focuses on a very basic set of AdWord metrics—specifically, impression shares, clicks, CTR, and video watch statistics.¹⁸

A campaign receives an impression every time an ad appears in the search page on Google or the Google network. Impression shares equal the percentage of impressions that an ad receives divided by the total possible number of impressions that the ads could get based on a user’s search behavior.¹⁹ Impression eligibility is based on a variety of factors, such as the quality score of the ad, the bid to have the ad placed, and other targeting settings (Google, undated).

Results for total number of impressions, clicks, CTR and impression share are presented in Table 1. As the table illustrates, the violent far-right campaign achieved a much higher number of total impressions, which likely indicates that, in the United States, this group is larger than the number of violent jihadist extremist sympathizers. The numbers suggest that approximately 2–3 percent of those exposed to Redirect ads clicked on the links, with average watch times ranging from 10 to 59 seconds. Results for impression shares varied significantly between the

two campaigns, with the violent jihadist campaign achieving more than 90 percent impression shares, compared with 51.23 percent for the violent far-right campaign. The impression share for the entire campaign was 55.29 percent.

What are the implications of these results? Moonshot sought to achieve an impression share rate of 75 percent for the violent jihadist and violent far-right campaigns combined.²⁰ The campaign targeting the violent jihadi population far exceeded this figure with an impression share of more than 90 percent. However, the total and combined impression share is weighed down by the violent far-right campaign, which has an impression share of slightly more than 50 percent and nearly five times as many total impressions as the violent jihadi campaign. Several factors complicated this for the violent far-right campaign. Search impression share is dictated by the advertisement’s quality score. In cases where the audience is large, the size of the competition can affect the quality score. This number also could be affected by how well the ad text directly relates to a user’s search term. Google company policy also played a role because the campaign was temporarily terminated several times when Google, apparently in an effort to limit advertising on racist search terms following the Unite the Right rally, inadvertently and temporarily suspended the Redirect Campaign. This required Moonshot to rebuild the campaign.²¹

To understand the value of the CTR result of 3.19 percent and 2.22 percent for the violent jihadist and violent far-right campaigns, respectively, it is helpful to compare these figures with those of other campaigns. For example, according to the social media analytics outfit WordStream, the average CTR across industries for AdWord campaigns is 3.17 percent (WordStream, undated).

TABLE 1
Impressions, Clicks, and Click-Through Rate

Engagement	Violent Jihadist	Violent Far Right	Combined Campaigns
Impressions	36,296	179,925	216,221
Clicks	1,158	4,010	5,168
CTR	3.19%	2.22%	2.39%
Impression share	91.13%	51.23%	55.29%

NOTE: CTR is the ratio of impressions (number who are shown an ad) and clicks (the number who click the ad).

The English-language Redirect Method campaign achieved a 3.1-percent CTR, which is obviously comparable to the violent jihadist campaign result (Redirect Method, undated). It should also be noted that WordStream’s average for advertising campaigns is largely based on Google ads that link to information that users want to access and consume. In contrast, the links for the Redirect Method, although neutrally titled, provide users access to information they do not want and that disagrees with their world view. In this sense, the reported CTRs for Redirect Method appear quite strong.

The most successful video for the violent jihadist campaign was Truth for Jihad (Tables 2 and 3). According to Moonshot CVE, this playlist addressed “the main narrative put forward by most jihadist groups, including ISIS, which dictates that Islam is under threat and that Muslims have a duty to protect their religion and other fellow Muslims across the world” (Moonshot CVE, 2017, p. 99). For the violent far-right campaign, the Fight for Your Culture playlist was viewed most frequently and it sought to debunk the message pushed by white supremacist and neo-Nazi groups regarding the duty to protect the white race against other, inferior races (Moonshot

CVE, 2017, p. 98). This was viewed for an average of 54 seconds by more than 3,000 individuals.

Recommendations for Redirect Method Evaluations

To what extent can the Redirect Method employ these evaluation strategies? Many of the survey evaluation strategies reviewed would likely be more easily implemented in “upstream” campaigns that seek to target nonradicalized youth or youth who are at risk of radicalization but not yet enamored with its cause. Such programs could, for example, seek to raise awareness, promote early detection, elicit participation in crowdsourced communication campaigns, or inoculate targeted individuals against future radicalization. Many of these campaigns could fairly easily employ testing strategies that assess the effect of content on knowledge, attitudes, and behavior and use those data to gain critical feedback on campaign content.

The Redirect Method poses a unique problem: It targets a potentially highly radicalized population that would in theory be reluctant to provide honest feedback to a counterradical program. Redirect

TABLE 2
Playlist Watch Time and Views for Violent Jihadist Intervention

Title	Watch Time (mins)	Views	Average Watch Time per View (min:sec)
The Truth About Jihad	589	768	0:46
Dangers to Muslims in the West	75	128	0:35
Sacrifice and Reward	44	121	0:21
Muslim Women in the West	38	59	0:38
Does the West Hate Islam?	7	43	0:10

TABLE 3
Playlist Watch Time and Views for Violent Far-Right Intervention

Title	Watch Time (mins)	Views	Average Watch Time per View (min:sec)
Fight for Your Culture?	2,936	3,233	0:54
White Genocide—Fact or Fiction?	1,286	1,303	0:59
Women Safe in the USA?	276	392	0:42
Religion of Peace?	112	194	0:35
Can We Trust the Government?	99	142	0:42

planners understandably have resisted connecting survey links to the campaign because of concerns that this would render the results less authentic and impactful. This leads to several evaluation options, which are presented here.

Process Evaluation with Former Extremists

At the most basic level, a process evaluation of search page advertisements and video content can be conducted with a sample of former extremists. Redirect Method planners consulted with a small cadre of former extremists (colloquially known as “formers”) to develop the list of targeted search terms. Formers have also been enlisted at an informal level to review “every piece of video.”²²

A systematically conducted evaluation with formers could develop insights into two key components of the Redirect Method. First, planners can seek to estimate audience perceptions of the videos and assess whether the content is entertaining, accurate, and potentially persuasive.²³ A process evaluation with former extremists could easily be conducted prior to campaign launch, and the results could not only feed into program development but also help validate campaign effectiveness.²⁴ Second, researchers could also examine the content of the Google search result ads, otherwise known as advertisement copy. It was noted that the violent far-right campaign achieved a slightly lower CTR (2.22 percent) than the violent jihadist campaign (3.19 percent). Researchers could identify ways to improve the advertisement copy and thus increase the CTR. A-B studies could also be employed to compare and ultimately select the best advertisement copy.

Closed Evaluation Design

In addition, it might be feasible to test the impact of the Redirect Method in a closed evaluation design, which would prevent the provision of survey content to viewers of the live campaign. For example, Facebook advertisements could be used to target and recruit panel participants. While it would not make sense to target highly extreme samples, it might be possible to identify audiences sympathetic to such efforts. Surveys can test impact. In addition, to mask the potential purpose of the testing, Redirect videos

could be nested within a group of otherwise irrelevant videos and randomization survey techniques—for example, a list experiment can subtly test whether the videos produced any sort of short-term or long-term impact.²⁵

Call to Action

Moonshot CVE is considering future iterations of the Redirect Method that would elicit “call to action” behavior. Part of this could include Moonshot directly reaching out to target audiences and attempting to engage in a one-on-one conversation with extremists. This approach might be similar in style to Frenett and Dow (2014) and Davey et al. (2018)—studies in which former extremists used Facebook as the vehicle to identify and connect with extremist users. In this case, planners can assess the audience response to this outreach.²⁶ It also might be possible for the campaign to ask users to click on an external link or to contact a helpline phone number to receive more information or to gain help in leaving their extremist circles. The number of links or the number and character of the phone calls could be an indicator of success.

Additional Evaluation Designs

Finally, it is worth noting one additional approach to evaluating the Redirect Method, which involves drawing on methods that do not involve user surveys. We previously noted that many researchers directly analyze YouTube and other video comments submitted to the intervention site. However, a new method for analysis draws on a much larger body of comments. This method, developed in collaboration with the Defense Advanced Research Projects Agency, involves developing a sample of users who post comments on the Redirect Method’s YouTube page, then pulling the comments that those users post to videos watched before and after watching Redirect Method videos. The question posed to researchers is whether the rhetoric of those individual comments changes, evincing either less or possibly more extremist content. Newly developed methodologies in text analytics provide a mechanism to analyze the content of such comments at scale (Coppersmith, Dredze, and Harman, 2014; Chaudhury et al., 2016).

The initial study that applied this methodology to the Redirect Method was never published, so it is impossible to provide an assessment on the study's results or empirical merits.²⁷ Such an evaluation would be subject to the same limitations as those already highlighted, specifically that those commenting on Redirect videos (and thus the sample size for the analysis) would be a self-selected minority of all of those exposed to the videos. However, such an analysis could be valuable in identifying whether a minority of users benefit from exposure to the Redirect Method videos.

Conclusion

In conclusion, periodic and rigorous evaluations are critical to ensure that online CVE campaigns are effective and that limited resources are properly allocated. Evaluations conducted for many such online campaigns do not evaluate the campaign's impact on target audience attitudes and or behaviors.

While impact evaluations can be complicated, this report reviewed several potential approaches,

drawing on a brief review of public health literature. Researchers must first choose whether to evaluate impact via an open or closed campaign. Surveys provide one evaluation option for gauging audience reactions, but care must be taken in how such surveys are constructed and how participants are recruited. Calls to action and analysis of comments can also provide unique approaches to measurement.

A limited evaluation of the Redirect Method process variables suggests that the implementers are able to use advertisements linking to counterextremist videos to effectively expose individuals searching for violent jihadist or violent far-right content to content that offered alternative narratives. Users clicked on these ads at a rate on par with industry standards. But as is the case with other CVE evaluations, this partial evaluation did not assess the impact of the video content on user attitudes or behavior.

The potentially highly radical nature of the Redirect Method's target audience makes evaluation of the campaign particularly complicated and thus might necessitate the recruitment of former extremists to help gauge audience response. Alternatively, it might be advisable to analyze user comments to understand how a subsample of users respond to the content.

TABLE A.1
Aspects of Campaign Studies

Campaign	Campaign Goals	SNS Used	Reach	Engagement	Impact	Other Comments
“One to One” pilot study: Former extremists sent direct messages to individuals whose Facebook activity suggested they supported white supremacist or Islamic extremist causes (Frenett and Dow, 2014)	To test the feasibility of an online campaign that seeks to dissuade online extremists	Facebook	<ul style="list-style-type: none"> Messages were sent to 76 extremists 	<ul style="list-style-type: none"> More than 60% of messages were read 59% responded directly or shifted behavior (e.g., changed their policy settings) Of those who did respond, 60% sustained engagement Casual, sentimental, and reflective tones performed best (as opposed to antagonistic or meditative tones) 	Not assessed	
Follow-up to the “One to One” pilot study: Former extremists engaged with more than 800 individuals on Facebook who identified as Islamist or right-wing extremists (Davey, Birdwell, and Skellett, 2018)	To test the feasibility of using direct online engagement with extremists to support their exit from extremism and hate	Facebook	<ul style="list-style-type: none"> Messages were sent to 814 extremists 	<ul style="list-style-type: none"> Initial response rate: One in five contacted individuals responded Sustained engagement rate: 71% of responding Islamist candidates; 65% of responding right-wing candidates engaged in sustained conversations 	Referred to as an indication of “potential positive impact,” 10% of sustained conversations suggested that the program had a positive impact, as evidenced by expressing interest in taking the conversation offline, indicating that the conversation might have changed attitudes or beliefs or that the conversation had a positive effect on negative online behaviors	
Assessed how counterextremist speech on Facebook is produced and shared (Bartlett and Krasodonski-Jones; 2015)	Separately examine counterextremist Facebook page accounts in France, India, Indonesia, Morocco, Tunisia, and United Kingdom	Facebook	<ul style="list-style-type: none"> Not examined 	<ul style="list-style-type: none"> Analyzed 1,000 randomly selected comments from right-wing posts and 1,000 comments on countercontent pages 	Not assessed	

Table A.1—Continued

Campaign	Campaign Goals	SNS Used	Reach	Engagement	Impact	Other Comments
“Say no to terrorism” campaign produced 13 videos targeting extremism in Saudi Arabia and posted them to YouTube (Al-Rawi, 2013)	Address extremism in Saudi Arabia	YouTube and Facebook	<ul style="list-style-type: none"> 1,348,791 YouTube views 	<ul style="list-style-type: none"> Analyzed 281 comments: 60% negative tone; 20% positive; 18% neutral/irrelevant Analysis suggests some audience suspicion regarding origin and intent of the campaign 	Not assessed	
“Extreme Dialogue” series of short, emotive films told the stories of individuals affected by violent extremism (Reynolds and Tuck, 2016)	Counterextremist narratives, dispel extremist myths, and encourage empathy with and understanding of “others”	YouTube, Facebook advertising, Twitter	<ul style="list-style-type: none"> YouTube: 50,673 videos watched Facebook: 362,500 users received a post; 1,835 clicks Viewer retention: Average of 37–65% of videos watched on average 	<ul style="list-style-type: none"> YouTube: 55 shares, 22 comments Facebook: Likes, shares or comments from 916 users 	Not assessed	Analysis of audience demographics suggests that YouTube generated a more age-appropriate audience
“Average Mohamed” campaign used five animated videos to “counter the ideology of Islamist extremist groups” (Reynolds and Tuck, 2016)	Empower young Muslims through counterideology messages and discourage them from joining Islamist extremist groups	Twitter, YouTube, Facebook	<ul style="list-style-type: none"> Facebook: 456,113 users received a post; clicks not reported 	<ul style="list-style-type: none"> Compared engagement/reach, which was highest on Twitter (1:14) compared with YouTube (1:304) and Facebook (1:141) 305 comments; 66% supportive; the rest were “negative” or “unrelated” 	Not assessed	

Table A.1—Continued

Campaign	Campaign Goals	SNS Used	Reach	Engagement	Impact	Other Comments
ExitUSA outreach campaign run by Life After Hate group that uses four videos highlighting experience of former extremists (Reynolds and Tuck, 2016)	To discredit far-right extremists and “sow seeds of doubt” in their members	Twitter, YouTube, Facebook	<ul style="list-style-type: none"> Facebook: 212,051 users received a post; clicks not reported 70% of viewers were male, 20% were female. 	<ul style="list-style-type: none"> Facebook: 2,127 total engagements; YouTube: 42 total engagements Twitter: 1,692; Achieved best engagement to impression ratio (1:123) Comments: 141; 32% supportive; 14% misunderstood, 18% negative, engagements across all platforms, with 2,127 from Facebook. 	Sustained engagement analysis showed that 8 users reached out to discuss deradicalization	
Harakat-ut-Taleem campaign launched six videos that use testimonials and dramatization (Silverman et al., 2016)	Dissuade young people from joining the Taliban in Pakistan and show western young people the destructive nature of violent extremist groups through their personal stories	Facebook	<ul style="list-style-type: none"> Facebook: 88,351; clicks not reported Viewer retention: average of 50 and 68% of videos watched 	<ul style="list-style-type: none"> Facebook total engagements: 646 YouTube total engagements: 3 Twitter total engagements: 4,814 		Urdu-speaking former violent extremist ensured that content was emotive and true to the situation in Pakistan

Notes

¹This control over exposure is particularly valuable because social media users typically choose the content they will consume, which can create inherent self-selection biases in typical cross-sectional designs that seek to compare users who were exposed to the campaign with those who were not. In addition, the control over exposure addresses a fundamental problem with using pre-post surveys to assess impact in open campaigns: The rapid flow of information and the often fleeting nature of watching a 30-second clip can reduce the accuracy of exposure recall (Andersen, de Vreese, and Albaek, 2016; Niederdeppe, 2016; Niederdeppe, 2014; de Vreese and Neijens, 2016). If participants in a cross-sectional survey design cannot accurately remember whether they were exposed to the relevant social media content, it becomes increasingly difficult to assess impact accurately.

²As Chung (2016) notes,

Social media platforms are communication tools particularly suitable for process evaluation . . . [S]ocial media platforms provide rich data on the reach of a campaign messages and the level of user engagement that constitute core questions for process evaluation.

³Unfortunately, it is not possible to identify audience demographics based on basic reach data. It is possible to determine demographics of audiences that do engage (like, share, comment) with the content. This is a significant limitation of the social media analytics because identifying whether a campaign's message reaches the intended target audience is key objective of process evaluations.

⁴We interviewed several experts, including two professionals in commercial marketing campaigns and three professionals in CVE evaluations.

⁵See Andersen, de Vreese, and Albaek, 2016; Niederdeppe, 2016; Niederdeppe, 2014; and de Vreese and Neijens, 2016.

⁶The manners in which participants can be exposed to social media content can vary. For example, participants can be asked to follow a specific Facebook page, which then produces content and disseminates it to followers' profiles. Another example would be sending video links to participants via email and asking them to click through and watch the provided content. Questionnaires, delivered over time, can be used to assess attitudinal changes.

⁷This control over exposure is particularly valuable because social media users typically choose which content they will consume, which can create inherent self-selection biases in typical cross-sectional designs that seek to compare users who were exposed to the campaign with those who were not. It also mitigates against the central challenge in cross-sectional survey assessments in open campaigns: inaccurate recall of message exposure (Andersen, de Vreese, and Albaek, 2016; Niederdeppe, 2016; Niederdeppe, 2014; de Vreese and Neijens, 2016).

⁸Bartlett, Birdwell, and Littler (2011) effectively conducted Facebook surveys with followers of far-right populist groups in Europe.

⁹Another study used street intercept surveys to assess the success of a social media (and other display media) program focused on diabetes prevention, nutrition, and fitness that targeted residents in Central Brooklyn and East New York.

¹⁰One study (Ramo et al., 2015) used Facebook to recruit young adult smokers and randomly assigned participants to either a private Facebook group, which served as a platform to specially serve the program's content, or a control smoking cessation website. The researchers then disseminated pre-post surveys to help assess impact. Young (2014) used Facebook to recruit at least part of that study's sample of men who have sex with men. (Others were recruited offline.) Participants were randomly assigned either to active intervention or control Facebook groups.

¹¹To address concerns about disinformation and election manipulation, Facebook has made it more difficult to use some audience demographic characteristics for its targeted advertisements (Cohen, 2018). But which is the best platform to recruit participants? Antoun et al. (2016) studied this issue, comparing online recruitment strategies used on Craigslist, Google AdWords, Facebook and Amazon Turk. Their findings indicated that methods that "pull" in participants who are already looking for work on Mechanical Turk and Craigslist tend to be more cost-efficient and those participants are more likely to complete survey panels. In contrast, methods that "push out" a recruiting ad to online users (Google AdWords and Facebook) tend to generate more demographically diverse audiences.

¹²An example of the two kinds of endorsement experiment questions would be (1) "Some states have raised the highway driving speed from 65 to 75 MPH [miles per hour]. How much do you support such plans?" and (2) "Some states have raised the highway driving speed from 65 to 75 MPH. The Klu Klux Klan have voiced support for these new laws. How much do you support such plans?"

¹³Blair, Imai, and Lyall (2014), for example, conducted a list experiment study in Afghanistan. They asked participants to read a list of names of different groups and then identify how many of those groups and individuals they broadly support. Participants were instructed not to name the groups but simply count the number. The list included nonsensitive items, such as the Karzai government, National Solidarity Program, and Local Farmers. The treatment group received the same list as the control group but with the additional sensitive item Foreign Forces, which refers to U.S. and international troops. Authors then statistically analyzed the data to derive an estimate of the population's support for U.S. and international troop presence in Afghanistan.

¹⁴Interview with marketing expert, March 20, 2017.

¹⁵The concept for this pilot program dates back to 2011 when Jigsaw, a subsidiary of Google's parent company Alphabet Inc., hosted the Summit Against Violent Extremism in Dublin that convened a varied consortium of former gang members, neo-Nazis, religious extremists, and survivors of extremist attacks. The conference highlighted the potential role that former extremists could play in countering violent extremism, and Jigsaw and the Gen Next Foundation launched a series of initiatives. Together with London's ISD, Jigsaw and Gen Next formed Against Violent Extremism, a global network of former extremists and survivors that work together to counter all forms of violent extremism. In addition, Jigsaw began helping the creator of Abdullah X, a counterextremist cartoon, to develop and fine-tune videos and effectively use Google's AdWords advertising technology. This helped provide valuable insights into the concept that advertising can help capture a targeted audience for CVE programs and it was recognized that this enterprise could be scaled by identifying a broader range of underlying extremist narratives, developing content to debunk those narratives, and using such technologies as AdWords to link that content to online searches (interviews with officials from Gen Next Foundation, May 4 and 18, 2018; interviews with officials from Moonshot CVE, December 12, 2017, and April 11, 2018; interviews with officials from Jigsaw, June 6, 2018).

¹⁶In 2017, for example, right-wing extremists conducted 37 violent attacks compared with seven committed by Islamic extremists (Romero, 2018). Data by the New America Foundation, for example, show that between the terrorist attacks of September 11, 2001, and 2015, right-wing and other nonjihadist extremists killed almost twice as many people as Islamic jihadists did (Shane, 2015).

¹⁷The violent jihadist campaigns took place between July 13 and November 15, 2018, and the violent far-right campaign took place from June 23 to September 15, 2018.

¹⁸The campaign produced very few comments (15 total for the violent far-right campaign and eight for the violent jihadist campaign). We did not analyze these data other than to confirm that none of the commenters reached out to program implementers for assistance. We also did not analyze other likes or shares, reasoning that the videos, which seek to confront deeply held radical opinions in the target audience, were unlikely to receive such responses.

¹⁹Impression share = impressions / total eligible impressions (Google, undated).

²⁰Interview with Moonshot CVE representative, December 7, 2017.

²¹Moonshot CVE specifically noted several factors that could have complicated its violent far-right campaign but not the violent jihadist one. Specifically, they observed the following four factors (Moonshot, email correspondence with author, June 14, 2018):

- For the violent far-right campaign, it is possible that, with some corporations bidding against the same search terms as the Moonshot CVE effort and using potentially larger budgets, Moonshot might have lost out on the AdWords "bidding war."

- Relatedly, the average cost per click on the violent far-right campaign was high in the first few days of the campaign and Moonshot had to reduce its bids to avoid overexpending on the campaign.
- Quality score can affect search impression share, and the quality score for the Redirect Campaign might have been relatively low because the content served by the Redirect Method is not necessarily the content that its target audience is seeking.
- The relevance of a landing page can have an impact on the search impression share. Certain techniques (for example, making the landing page more relevant by including keywords that the campaign is bidding against) do not apply to the Redirect campaign because users are directed to YouTube rather than a conventional website.

In the future, better matching keywords that campaigns are bidding against with the text in the ads themselves will help improve search impression score.

²²Interviews with representatives from Moonshot CVE, December 12, 2017, and April 11, 2018.

²³Researchers should also pay close attention to audience reactions to video content. Many of the campaign's audience members stopped watching videos after 10–59 seconds. Some participants likely terminated once they realized the counter-extremist nature of the content, but user experience research might identify ways to extend viewer time.

²⁴Campaign planners have generally avoided conducting systematic evaluations with formers out of concern that formers do not represent the target audience of active extremists. In our view, former extremists can reflect on the extremist experience in ways that non-extremists cannot, even if that perspective is not perfect. In the absence of other available assessment methods, the perspective of formers is likely better than nothing.

²⁵One additional benefit of such a design is that it could be used to rule out any potential negative effects of the Redirect campaign. There is growing concern that some CVE campaigns produce unanticipated negative side effects—specifically, several studies have shown that at least a subset of the target audience has experienced an increase in radicalization. Why might some CVE campaigns increase radicalization? One theory is that media attacking strongly held beliefs can make those beliefs more resistant to argument. The media, for example, could spark an immediate defensive response or argument in the mind of an audience member that can reinforce and strengthen pre-existing radical views (see Darden, 2018, for a review). This could be a particular risk for the Redirect Method given the extremist views of its target audience,

²⁶As Davey, Birdwell, and Skellett (2018) used the term "potential positive impact," which they characterized as an online extremist expressing interest in taking the conversation offline, indicating that the conversation might have changed attitudes or had a positive effect on negative online behaviors.

²⁷Interview with Defense Advanced Research Projects Agency analyst, July 7, 2018.

References

- Al-Rawi, Ahmed K., "The Anti-Terrorist Advertising Campaigns in the Middle East," *Journal of International Communication*, Vol. 19, No. 2, 2013, pp. 182–195.
- Andersen, Kim, Claes de Vreese, and Eric Albaek, "Measuring Media Diet in a High-Choice Environment: Testing the List-Frequency Technique," *Communication Methods and Measures*, Vol. 10, No. 2–3, 2016, pp. 81–98.
- Antoun, Christopher, Chan Zhang, Frederick G. Conrad, and Michael F. Schober, "Comparisons of Online Recruitment Strategies for Convenience Samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk," *Field Methods*, Vol. 28, No. 3, 2016, pp. 231–246.
- Bartlett, Jamie, Jonathan Birdwell, and Mark Littler, "The Rise of Population in Europe Can Be Traced Through Online Behavior: The New Face of Digital Populism," *Demos*, 2011. As of October 4, 2018: https://www.demos.co.uk/files/Demos_OSIPOP_Book-web_03.pdf
- Bartlett, Jamie, and Alex Krasodonski-Jones, "Counter-Speech: Examining Content that Challenges Extremism Online," *Demos*, October 2015.
- Blair, Graeme, Kosuke Imai, and Jason Lyall, "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan," *American Journal of Political Science*, Vol. 58, No. 4, October 2014, pp. 1043–1063.
- Chaudhury, Munmun De, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar, "Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2098–2110.
- Chung, J. E., "A Smoking Cessation Campaign on Twitter: Understanding the Use of Twitter and Identifying Major Players in a Health Campaign," *Journal of Health Communication*, Vol. 21, No. 5, 2016, pp. 517–526.
- Clayforth, C., S. Pettigrew, K. Mooney, I. Lansdorp-Vogelaar, M. Rosenberg, and T. Slevin, "A Cost-Effectiveness Analysis of Online, Radio and Print Tobacco Control Advertisements Targeting 25-39 Year-Old Males," *Australian and New Zealand Journal of Public Health*, Vol. 38, No. 3, 2014, pp. 270–274.
- Cohen, David, "Facebook Will No Longer Allow Exclusion of Certain Groups Via Its Ad Targeting Tools," *Adweek*, July 27, 2018. As of October 4, 2018: <https://www.adweek.com/digital/facebook-will-no-longer-allow-exclusion-of-certain-groups-via-its-ad-targeting-tools/>
- Coppersmith, Glen, Mark Dredze, and Craig Harman, "Quantifying Mental Health Signals in Twitter," *Workshop on Computational Linguistics and Clinical Psychology: Linguistic Signal to Clinical Reality*, 2014, pp. 51–60.
- Darden, Jessica Trisko, "Compounding Violent Extremism? When Efforts to Prevent Violence Backfire," *War on the Rocks*, June 6, 2018.
- Davey, Jacob, Jonathan Birdwell, and Rebecca Skellett, *Counter Conversations: A Model for Direct Engagement with Individuals Showing Signs of Radicalization Online*, London: Institute for Strategic Dialogue, 2018.
- de Vreese, Claes H., and Peter Neijens, "Measuring Media Exposure in a Changing Communications Environment," *Communication Methods and Measures*, Vol. 10, No. 2–3, 2016.
- Fallin, A., T. B. Neilands, J. W. Jordan, and P. M. Ling, "Social Branding to Decrease Lesbian, Gay, Bisexual, and Transgender Young Adult Smoking," *Nicotine & Tobacco Research*, Vol. 17, No. 8, 2015, pp. 983–989.
- Frenett, Ross, and Moli Dow, *One to One Interventions: A Pilot CVE Methodology*, London: Institute for Strategic Dialogue and Curtian University, 2014. As of October 10, 2018: http://www.strategicdialogue.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf
- Google, "AdWords Help: About Impression Share," webpage, undated. As of July 13, 2018: <https://support.google.com/adwords/answer/2497703?hl=en&co=ADWORDS.IsAWNCustomer%3Dfalse>
- Greenberg, Andy, "Google's Clever Plan to Stop Aspiring ISIS Recruits," *Wired*, September 7, 2016. As of October 5, 2018: <https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits>
- Helmus, Todd C., Rajeev Ramchand, Sina Beaghley, David Stebbins, Amanda Kadlec, Michael A. Brown, Aaron Kofner, and Joie D. Acosta, *RAND Program Evaluation Toolkit for Countering Violent Extremism*, Santa Monica, Calif.: RAND Corporation, TL-243-DHS, 2017. As of June 29, 2017: <https://www.rand.org/pubs/tools/TL243.html>
- Lim, M. S., C. J. Wright, E. R. Carrotte, and A. E. Pedrana, "Reach, Engagement, and Effectiveness: A Systematic Review of Evaluation Methodologies Used in Health Promotion Via Social Networking Sites," *Health Promotion Journal of Australia*, Vol. 27, No. 3, 2016, pp. 187–197.
- Moonshot CVE, email correspondence with author, June 14, 2018.
- Moonshot CVE, *The Redirect Method: USA, Final Report*, unpublished briefing, December 2017.
- Mowery, A., P. Riedesel, M. Dreher, B. A. Schillo and J. E. Saul, "Using Online Message Testing to Evaluate TV Ads, Select Effective Messaging, and Improve Public Health Campaigns," *Social Marketing Quarterly*, Vol. 22, No. 33, 2016, pp. 179–199.
- Nanes, Mathew, and Bryony Lau, *Surveys and Countering Violent Extremism*, San Francisco, Calif: The Asia Foundation, 2018.
- Niederdeppe, J., "Conceptual, Empirical, and Practical Issues in Developing Valid Measures of Public Communication Campaign Exposure," *Communication Methods and Measures*, Vol. 8, No. 2, 2014, pp. 138–161.
- Niederdeppe, Jeff, "Meeting the Challenge of Measuring Communication Exposure in the Digital Age," *Communication Methods and Measures*, Vol. 10, No. 2–3, 2016, pp. 170–172.
- Oksanen, Atte, David Garcia, Anu Sirola, Mattie Näsi, Markus Kaakinen, Teo Keippip, and Pekka Räsänen, "Pro-Anorexia and Anti-Pro-Anorexia Videos on YouTube: Sentiment Analysis of User Responses," *Journal of Medical Internet Research*, Vol. 17, No. 11, 2015.
- Ramo, D. E., J. Thrul, K. L. Delucchi, P. M. Ling, S. M. Hall, and J. J. Prochaska, "The Tobacco Status Project (TSP): Study Protocol for a Randomized Controlled Trial of a Facebook Smoking Cessation Intervention for Young Adults," *BMC Public Health*, Vol. 15, No. 1, 2015, pp. 1–9.
- Redirect Method, homepage, undated. As of June 4, 2018: <https://redirectmethod.org>

Reynolds, Louis, and Henry Tuck, *The Counter-Narrative Monitoring & Evaluation Handbook*, London: Institute for Strategic Dialogue, 2016.

Robertson, Campbell, Christopher Mele, and Sabrina Tavernise, "11 Killed in Synagogue Massacre: Suspect Charged with 29 Counts," *New York Times*, October 27, 2018. As of November 10, 2018:
<https://www.nytimes.com/2018/10/27/us/active-shooter-pittsburgh-synagogue-shooting.html>

Romero, Luiz, "Terrorism Is Surging in the U.S., Fueled by Right-Wing Ideologies," *Quartz*, August 17, 2018. As of October 10, 2018:
<https://qz.com/1355874/terrorism-is-surging-in-the-us-fueled-by-right-wing-extremists>

Shane, Scott, "Homegrown Extremists Tied to Deadlier Toll Than Jihadists in U.S. since 9/11," *New York Times*, June 24, 2015. As of October 10, 2018:
https://www.nytimes.com/2015/06/25/us/tally-of-attacks-in-us-challenges-perceptions-of-top-terror-threat.html?_r=0

Silverman, Tanya, Christopher J. Stewart, Sahed Amanullah, and Jonathan Birdwell, *The Impact of Counter-Narratives: Insights from a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact*, London: Institute for Strategic Dialogue, 2016.

Sobhani, Parinaz, Saif M. Mohammad, and Svetlana Kiritchenko, "Detecting Stance in Tweets and Analyzing Its Interaction with Sentiment," *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (*SEM 2016)*, Berlin, Germany, August 11–12, 2016, pp. 159–169.

WordStream, "Average CTR (Click-Through Rate): Learn How Your CTR Compares," webpage, undated. As of June 14, 2018:
<https://www.wordstream.com/average-ctr>

Young, Sean D., Ian Holloway, Devan Jaganath, Eric Rice, Drew Westmoreland, and Thomas Coates, "Project HOPE: Online Social Network Changes in an HIV Prevention Randomized Controlled Trial for African American and Latino Men Who Have Sex with Men," *American Journal of Public Health*, Vol. 104, No. 9, September 1, 2014, pp. 1707–1712.

About This Report

The number of programs dedicated to countering violent extremism (CVE) has grown in recent years, yet a fundamental gap remains in the understanding of the effectiveness of such programs. This is particularly the case for CVE campaigns, which are increasingly conducted in the online space. The goal of this report is to help CVE campaign planners better evaluate the impact of online efforts. It reviews prior assessments of online CVE campaigns, provides recommendations for future assessments, and provides a case study of one particular CVE campaign—the Redirect Method. The methodology and results should help the Redirect Method and other CVE campaign planners and evaluators better assess the impact of their campaigns.

Many people contributed to the completion of this report. We are especially grateful to the Gen Next Foundation for its unique vision in supporting this study and in so doing, supporting empirical evaluations of not only the Redirect Method but also online counter violent extremism campaigns more broadly. We are further grateful that the Redirect Method’s managers and developers for their willingness to make program data available for assessment. Finally, we offer many thanks to Elizabeth Bodine-Baron of the RAND Corporation and Louis Reynolds, formerly of the Institute for Strategic Dialogue, for their considered critique and observations in their reviews of this work. Any and all errors in this report are the sole responsibility of the authors.

This research was sponsored by the Gen Next Foundation and the RAND Center for Global Risk and Security (CGRS) and was conducted within the International Security and Defense Policy (ISDP) Center of the RAND Corporation’s National Security Research Division (NSRD).

NSRD conducts research and analysis for the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the defense agencies, the Navy, the Marine Corps, the U.S. Coast Guard, the U.S. Intelligence Community, allied foreign governments, and foundations. For more information on the RAND ISDP Center, see www.rand.org/nsrd/ndri/centers/isdp or contact the director (contact information is provided on the webpage).



The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

For more information on this publication, visit www.rand.org/t/RR2813.

© Copyright 2018 RAND Corporation

www.rand.org