



WASHINGTON STATE UNIVERSITY
EXTENSION

A Synthesis and Annotated Bibliography on Stormwater Behavior Change Campaigns

Wisnu Sugiarto and Joseph Cook

School of Economic Sciences

Washington State University

FINAL REPORT

June 21, 2022



Please cite as “Sugiarto, W. and J. Cook. “A Synthesis and Annotated Bibliography of Stormwater Behavior Change Campaigns”. Washington State University Stormwater Center. June 2022.

1. Introduction

This report describes the results of a comprehensive review of peer-reviewed studies and “gray literature” reports that have evaluated behavior change campaigns in water quality or stormwater. It is part of a larger project on these campaigns for the [Stormwater Action Monitoring](#) (SAM) program, funded by permittees and administered by the Washington Department of Ecology. The title of the larger project is “Evaluating the effectiveness of stormwater education and outreach: permittee guidance for addressing challenges through behavior change”, though the project team and our Technical Advisory Committee agreed to narrow our focus to behavior change. By “behavior change”, we mean programs that seek to get people to change a concrete behavior, such as picking up pet waste or reducing fertilizer use, rather than other “education and outreach” programs that seek to communicate information or change attitudes or programs that provide stewardship and volunteer opportunities without clear quantitative measures.

Our charge from SAM for Task 3 was to “conduct a broad literature review of education and outreach (E&O) programs, organized by stormwater issue...”, including a “3-10 pg. draft review of the literature (both “gray” and peer-reviewed) on the effectiveness of E&O programs, organized by stormwater issue, including studies from within and outside Washington State.” The report was also to “evaluate the pros and cons and research quality (e.g. evaluation of methods and data) of each existing E&O evaluation study.”

This report identifies, evaluates, and summarizes existing studies on behavior change campaigns associated with water quality or stormwater management. The audience is behavior change professionals who work in the stormwater and water quality space. This document is organized as follows. First, we discuss our approach in reviewing the literature. We adopt a comprehensive, objective, and reproducible search strategy to capture all relevant studies. Second, we discuss our findings and identify the common trends through descriptive statistics. Third, we highlight evaluation methods that have been used in assessing the effectiveness of behavior change programs. We supplement the evaluation strategy discussions with methodological concerns to identify gaps that can be used to improve future evaluations. Finally, we present a table summary of all collected articles, which is available as an appendix. The articles are also searchable and downloadable on www.waterbehaviorchange.org.

2. Review Criteria

The first guiding principle in conducting our review was relevance. We defined relevance as program reports and published academic articles that evaluated measurable and direct behavior change programs for water quality or stormwater management purposes. Behavior change studies could be “quantitative” or “qualitative” in nature. “Direct” behavior change programs reach their target audience directly, so we excluded studies that focused only on creating awareness or providing stewardship opportunities. Although an awareness campaign can induce behavior change, evaluations of these programs typically focus on changes in awareness or attitudes. Similarly, stewardship programs ask volunteers in the community to help with water quality or stormwater projects (e.g. storm drain monitoring, tree planting), though the aim is again to build awareness. We also excluded programs which focused on surveying residents’ perceptions of water quality, assessing people’s knowledge on stormwater issues, and/or reporting only homeowners’ intent to change. Similarly, there are a number of case studies that discuss the implementation of a social marketing approach in water quality but that had very little or no information on how the program was evaluated. We excluded studies that did not discuss evaluation at all, or only briefly in passing. We included studies that explained how an evaluation was conducted, even if the report provided no detailed information on the evaluation results or the evaluation attempts did not yield any results.

The second principle guiding the search was breadth. We employed multiple approaches to collect articles on behavior change programs related to stormwater management. First, we searched for published academic articles and publicly-available reports on several keyword databases such as [Google Scholar](#), [GreenFile \(EBSCO\)](#), [Academic Search Complete](#), and [Science.gov](#). Second, we found two databases that already contained a collection of stormwater program-related surveys and evaluations. The NPS Outreach Toolbox on the US EPA’s website has a collection of [Surveys & Evaluations](#) from around the United States. In fact, many of our collected reports came from the US EPA’s site. We also included case studies found on the [Tools of Change](#) and [CBSM websites](#). These were primarily community programs from across North America, and they covered other environmental issues beyond water pollution. We also collected program reports from our Washington key informant interviews. Finally, we contacted stormwater professionals or environmental managers at every U.S. state agency, including the U.S. territories. The list of all state agency contacts could be found on the United States Environmental Protection Agency (EPA) page for [stormwater contacts](#). Some of the listed contacts were incorrect, had retired, or were no longer working in the position. Thus, we searched online to obtain the updated contact information.

Finally, we aimed for our literature searches to be replicable. We searched Google Scholar, GreenFile (EBSCO), Academic Search Complete, and Science.gov with the following sets of keywords: 1) “stormwater” “behavior change” AND “education and outreach”, 2) “stormwater” “behavior change”, and 3) “stormwater” AND “behavior change”. This yielded a large number of results, though many were broken links, duplicative studies or were not relevant. We scraped the data, which were the large number of search outputs, and conducted a blind-review process. Each author independently reviewed each collected title and determined if it was relevant to our literature review study. We discussed discrepancies in coding and agreed on which to proceed with. We next independently reviewed the study abstracts. Some abstracts clearly fit the relevance criteria, but not all studies or reports can be assessed from their abstracts. We reviewed all studies and that both authors coded for inclusion based on the abstracts. This blind-review process allowed us to systematically collect search results that meet the relevance criteria and eliminate the ones that do not.

3. Data recorded

We systematically recorded several pieces of information from each collected study and report so that we could provide descriptive summary statistics and evaluate research quality in a consistent and transparent way. The data recorded also allowed us to compile the spreadsheet database (Appendix #). From each study, we recorded the study year, location, targeted behavior, and target audience. We recorded whether the program specifically targeted under-represented groups, such as non-native English speakers or racial minority groups. We abstracted information on the implementation of the program, including whether the implementers used educational posters, distributed brochures and posters, conducted home visits, used financial incentives (including subsidized purchases or free services like soil testing) or held online or in-person workshops or demonstrations.

To help us classify evaluations, we created a series of dummy (0/1) variables that captured each of the following categories, each of which is described in more detail below: a) well-defined **target audience**, b) well-defined **target behavior**; c) validated or pretested data collection **instruments**; d) evaluation included a **comparison group**; e) evaluation includes **pre-intervention data**; f) evaluation uses **observational data** (rather than self-reported data); g) Evaluation measures **long-term** change (≥ 1 year); h) evaluation discusses possible **selection bias** in uptake among target audience; l) evaluation includes **water quality measurements**. Every study was examined based on the 9 categories. Each category would

yield 1 point if the study met the criteria for that category. For example, category (a) was assigned 1 point if the study had a well-defined target audience. Otherwise, it would be assigned 0 points. Thus, the maximum score each study could get was 9. Although the summative score does not distinguish between cases where the information is missing (e.g. no information on pretesting the instruments is discussed) versus cases where sufficient information is reported to allow us to definitely classify the study as “0” (e.g. the study reports that its outcome measure is self-reported behavior), this data is recorded separately in the database. We group studies with scores below 4 as “*fair*”, studies with scores between 4 and 6 as “*good*”, and scores of 7-9 as “*exemplary*”.

Target audience, target behavior: All the major social marketing frameworks share a focus on clearly identifying a specific, end state target behavior and a specific target audience. Although this is largely about implementation rather than evaluation, we include them in our summative measure of study quality as a rough indicator of how well the practitioner applied social marketing ideas. Each is separately worth one point.

Validated instruments: Many studies rely on surveys to collect information on self-reported behavior or attitudes. Good survey practice dictates that survey questions are tested or validated in advance with focus groups or pretests or both. Evaluators can also re-use survey questions used by others that have already been tested.

Baseline data: Collecting pre- and post-intervention data is useful to attribute behavior changes to the program. If a campaign only employed a post-intervention evaluation, it would not have any comparison data to measure the effect of an intervention on program participants. For example, suppose there was a workshop on educating homeowners on substitutes for pesticides, and participants were only asked to complete a post-event survey. The collected survey results suggested that 75% of workshop participants purchased pesticide substitutes after attending the workshop. While the workshop had been helpful in informing participants about the danger of pesticides, it is possible that some of the participants had already planned on using alternative options in the first place. In this case, the workshop primarily served as a refresher course to reaffirm participants to move away from using pesticides, and one cannot conclude that the campaign increased substitute purchases by 75% percentage points. **Comparison or control groups:** Having a control or comparison group helps to establish that the program caused the desired behavior change by isolating its effect from other possible explanations for changes in behavior.

It provides a comparison between people or areas (within the defined target audience) that were exposed to the social marketing campaign and people or areas (again within the defined audience) that were not exposed. As an example, suppose jurisdiction A implemented a city-wide campaign targeted at new pet owners to encourage them to pick up and dispose of pet waste properly. To assess the campaign's effectiveness, behavior among new pet owners in jurisdiction B could in theory also be observed. Alternatively, half of new pet owners in jurisdiction A might be randomly assigned to receive the campaign materials (e.g. a "scoop the poop" sticker and free dog bag dispenser), but follow up surveys on self-reported behavior would be sent to both groups. Note that in both cases changes in behavior in the comparison group help control for other factors that might have changed behavior over the same time period. Finding the resources to identify and survey control groups may, however, be challenging.

Observational vs. Self-reported data: While most studies rely on surveys to ask households to self-report behaviors, this approach suffers from well-known reporting biases. This is particularly true when respondents believe there is a "right" or "socially-appropriate" answer¹; they might be ashamed of answering otherwise and there is little cost to being untruthful. This is clearly a concern in many water quality or stormwater-related behaviors (e.g. picking up pet waste, not dumping chemicals down the drain). Social marketing approaches may even attempt to build a social norm, so they in effect train respondents what the "right" or socially appropriate answer is. Good survey techniques can help ameliorate but not eliminate social desirability bias. Observational studies are very often not possible or practical (i.e. observing whether someone dumps chemicals down their kitchen drain), but have been done in programs for pet waste, dumpster maintenance and others.

Selection bias: Here we record whether the study monitored or discussed differences in uptake *among the target audience*. The main concern is that bias in who takes up the campaign can skew predictions about how a program may scale or expand beyond in the same jurisdiction or a different one.

Example: Suppose a city runs a natural yard care workshop for seniors. Among seniors, the only people who attend are those who are already very environmental-minded (perhaps as measured

¹ For more information, including techniques to reduce social desirability bias, see https://en.wikipedia.org/wiki/Social-desirability_bias

by a survey). Twenty people participate and the team evaluates their lawn care self-reported behaviors before and after the workshop with a survey. Suppose that the workshop is associated with increases in the target behavior, and the team asks for funding to run another 5 workshops. But the next 5 workshops may have less or no effect, and it may not even be possible to get anyone to attend: the program already harvested the “low-hanging fruit”.

Selection bias affects validity. The same program as the example with the same target audience in a different city may not find the same effect if they had fewer environmentally-minded seniors. As in control groups, it can be challenging to find resources to observe the characteristics of *non-participants* (or the community at large) to make these sorts of comparisons, though evaluators can sometimes use existing recent community-wide surveys (for things like environmental attitudes) or census data (for gender, income, age, etc.).

Long-term: Behavior changes induced by a social marketing campaign may be short-lived, so we distinguish between evaluations that looked only at short-term effects versus those that attempted more longer-term evaluations. We (somewhat arbitrarily) chose one year (or longer) as our cutoff for “long-term”.

Water quality information: Our primary focus is on the question of whether a behavior change campaign changed behavior, not whether the campaign actually improved water quality. Even campaigns that successfully change behavior may not cause detectable changes in water quality, depending on the hydrology and pollutant loading. The frequency and timing of water quality monitoring (which could be run by another department or agency) may not sync with the timing of the campaign. Nevertheless, we wanted to highlight the few studies in the literature that did attempt to measure water quality outcomes.

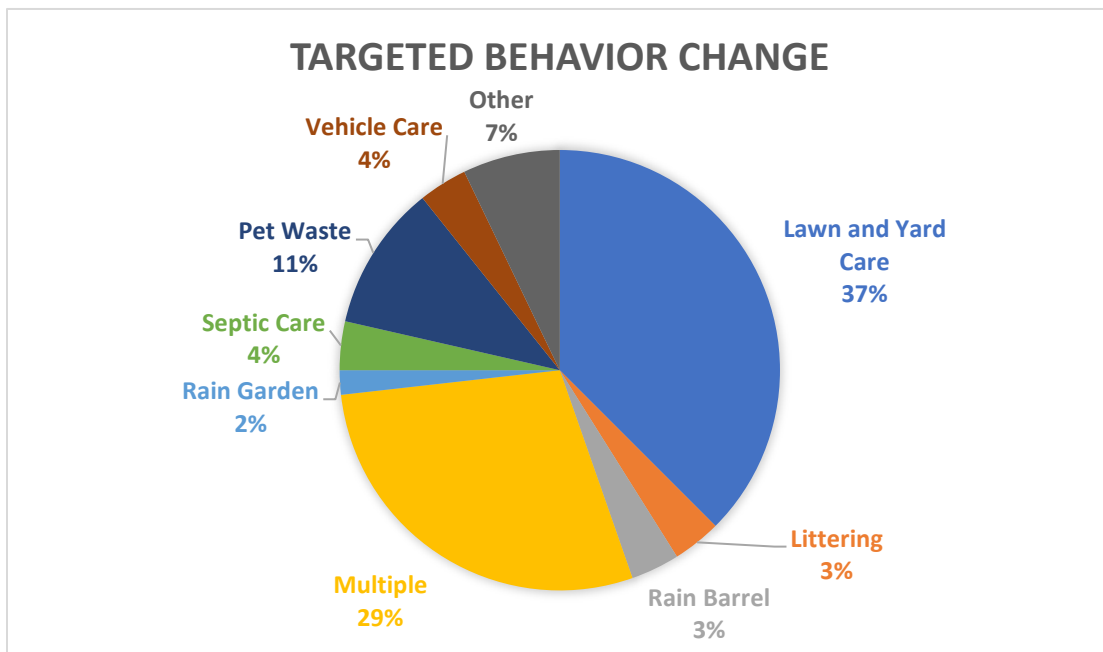
4. Descriptive Statistics on Studies

In total, we judged 56 documents to be relevant for review according to the criteria described in section 2. Fourteen were published in academic journals and the remaining 42 were program reports. One of the documents contained a report of eight case studies that discuss the impact of regulations and public education programs on reducing use of pesticides on residential property. In this case, information from the eight case studies were collected but the document containing the case studies was not counted to avoid redundancy. The oldest document was published in 1979 and the newest reports were released in 2021. Many municipalities involved an external consulting company to implement a behavior change program or conduct a program evaluation. Not surprisingly, there are also municipalities that involve a

university extension or research organization. For example, several Municipal Separate Storm Sewer Systems (MS4s) in Washington partnered with Washington State University Extension.

We categorized target behaviors into nine categories: 1) lawn and yard care practices, 2) rain barrel adoptions, 3) rain garden installations, 4) pet waste pick up and disposal, 5) litter pick up and disposal, 6) septic care, 7) vehicle care, 8) other, and 9) multiple (Figure 1). The lawn and yard care category included many practices, such as reducing the use of pesticides and fertilizers, using more environmental-friendly landscaping products, and recycling grass clippings. The vehicle care category included car washing, fixing oil leaks, and any programs targeting vehicle owners such as reduction in vehicle miles driven. The “other” category would also include a wide range of programs, such as care for street trees, proper disposal of used cooking oil and hazardous household wastes, and proper disposal of wastewater from carpet cleaning. The “multiple” category indicated programs that aim to change more than one behavior. For example, a campaign to maintain the health of the Nippersink Creek Watershed in Illinois wanted people to maintain septic systems properly and use phosphorus-free fertilizers.

Figure 1: Proportions of Targeted Behavior Change

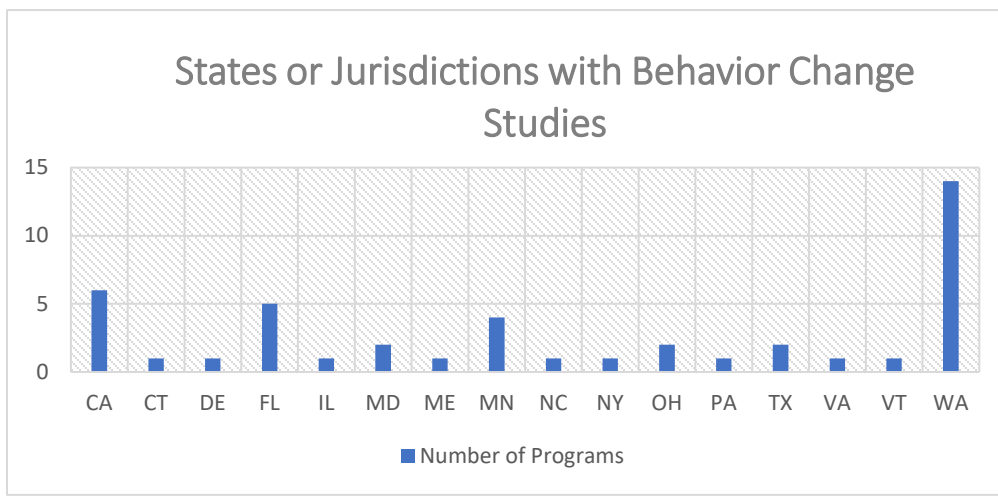


We found that programs related to lawn and yard care to be the most common; 37% (21 of 56) of the collected literature exclusively focused on lawn care practices to reduce the impact of stormwater runoff. Programs that employed multiple targeted behaviors were the next most common, with 16 studies. For example, the Think About Personal Pollution (TAPP) campaign in Tallahassee, Florida, aimed to change people’s behavior through switching to non-phosphorus fertilizers and picking up pet waste. Within these

16 studies, behaviors related to lawn care (29%) and pet waste disposal (17%) were the most common. Other relatively more common studies in the multiple group targeted rain garden or rain barrel adoption. The next most common type of exclusive behavior targeted (11% or 6 of 56) was picking up pet waste, followed by 4% (2 studies) concerned exclusively with septic care and another 4% (2 studies) focused on issues associated with vehicle care.

Geographically, many reports came from organizations in areas near estuaries or lakes, such as municipalities in the Puget Sound (Washington), California, Minnesota, and states surrounding the Chesapeake Bay, such as Maryland and Delaware (see Table 1). We also located several studies from outside of the United States, including programs in Australia, Canada, and Denmark, but note that Figure 2 reports the distribution of collected behavior change studies in US states only. Programs may actually be more common in these areas because of the importance of large, shared waterbodies, politics, larger cities or because many early social marketing studies originated in the Pacific Northwest. It may also be that jurisdictions in those areas were more likely to publicly report their findings. This also reflects differences in regulatory environments and the number of regulated cities.

Figure 2: Distribution of Collected Behavior Change Studies in the United States



Only two of the studies examined interactions with regulations or fines. For example, the small towns of Hudson, St. Lazare, and Notre Dame de L’Ille (in Montreal, Quebec) regulated the use of pesticides on residential property (Kassirer and Wolnik, 2004). On May 6th 1991, Hudson became the first municipality in Canada to pass a by-law regulating the use of pesticides on residential property. They employed a

variety of techniques to help residents comply, including the use of home patrols to answer questions and provide advice on alternatives to pesticides. Offenders were warned through warning letters with a threat of escalating fines. The regulation enforcement was also supplemented with the municipalities' efforts to build pride in being pesticide-free areas. For example, a pesticide-free campaign logo was used on all city paperwork for several years in Notre Dame and St. Lazare. The authors found that only communities that supported a regulation with education or made a community agreement were successful in reducing pesticides by a large amount (51-90%). E&O programs without regulations were more popular, but none achieved more than a "low" (10-24%) reduction during the study period

An example of a community agreement is the Drastrup project in Denmark. This pesticide reduction project was managed by the Aalborg municipal government to prevent groundwater contamination and meet European Union drinking water standards. There is a groundwater recharge area near Aalborg, Frejlev, which had a population of 2,000 people. As part of the agreement, land was purchased and farmers had to relocate outside the catchment area. Residents were provided with information about the consequences of pesticide contamination through meetings, website, local media, and newsletter.

The second study involving the use of a regulation also took place in Canada. The Halifax Regional Municipality was the first Canada's larger communities to introduce a pesticide by-law. The implementation was phased in over two years, between April 2001 and April 2003. The municipality contracted Clean Nova Scotia, a local non-profit organization, to process applications for pesticide permits through home visits. These home visits allowed program staff to educate property owners on sustainable maintenance practices

5. Program Evaluation Assessment

In this section, we identify trends, common practices, and gaps in existing evaluations. The purpose is to inform how behavior change evaluations might be improved, not to criticize individual studies, authors, or jurisdictions. We recognize that implementing staff operate under real time and money constraints and vary in their expertise with social marketing or program evaluation.

All (100%) of our curated and evaluated studies (56) for this project had at least one well-defined target audience and at least one well-defined target behavior. Program managers were clear with who they wanted to persuade and what behavior they wanted to change. In almost half (26 of 56) of the collected studies, staff validated their data collection instruments. Common validation practices included pre-testing surveys before dissemination and incorporating focus groups to supplement survey findings. For example, the Mobile Contractor Illicit Discharge Education & Outreach program led by the City of

Wenatchee, Washington, had a pilot test in the City of Pullman, Washington, to validate their survey instrument.

Table 1. Summary of evaluation quality measures

Measure	Percent of studies (n=56)
Well-defined target audience	100%
Well-defined target behavior	100%
Validated instruments	46%
Collected baseline/pre-intervention data	63%
Collected data on control or comparison group	13%
Use observed rather than self-reported data	26%
Address possible selection bias	48%
Measure behavior change after 1 year	36%
Measure water quality	18%

In contrast, we found that very few (7 of 56, or 13%) mentioned the use of a control group. Sixty-three percent of studies (35 of 56) included baseline data to measure changes in behavior caused by the campaign. Approximately 73% (41 of 56) measured behavior solely as self-reported in surveys administered by mail, telephone, or the web surveys. The remainder used observations or sales data to measure changes before and after an intervention. (For example, please see “Community Program #2: Halifax Regional Municipality, Nova Scotia” and “Community Program #5: Chesapeake Bay, Pennsylvania (Harrisburg)” in the bibliography table.) Observation was more common in published journal articles: over half of them included observed behavior.

Thirty six percent (20 of 56) of studies measured “long-term” changes (one year or longer). A few studies were clearly and intentionally interested in measuring the long-term effects of on-going programs. For example, Snohomish County Natural Yard Care Education team deployed a long-term post-outreach survey that was administered five years after receiving education. Approximately half (27 of 56) of studies monitored or discussed potential issues associated with selection bias. These studies did not necessarily analytically correct for any selection bias concern, but at least acknowledged the issue. We found 18% (10 of 56) of studies included some types of water quality measurement. These were relatively more common among the published journal articles, where 36% (5 of 14) included water quality measurement.

Considering the nine categories that we used to help us with measuring program evaluation quality, our review suggests that more than half (37 of 56, or 66%) of the collected behavior change studies are in

the “good” category, scoring between 4 and 6. One quarter (14 of 56, or 25%) of studies would be ranked as “fair”, scoring less than 4. No study achieved a “perfect” score of 9, but 9% (5 of 56) of studies would be categorized as “exemplary”, scoring 7 or 8.

Do these programs “work”? Evaluation results from multiple studies report that behavior change campaigns work in influencing people’s behavior to adopt stormwater runoff best management practices (BMPs). Among the collected studies, 63% (35 of 56) collected baseline data and provided some information on their campaign results. Nine of the 35 studies also collected observational or monitoring data. Outcomes of the behavior change campaigns are generally positive with 57% (18) of the 35 studies reporting successes in persuading people to adopt more environmental-friendly behavior. Forty three percent (15 of 35) of the studies reported what could be described as mixed results, where campaigns had positive and negative outcomes or where positive outcomes were attenuated in some way. For example, Thurston County Washington conducted a long-term evaluation in 2020 of their 2014 behavior change campaign, the GoGreen lawn care program. The evaluation found that the percentage of respondents who said that they used a weed-and-feed product fell from 65% at baseline to 12% in the 6-12 months after training. The reduction in weed-and-feed behavior remained in a follow-up survey six years post-training, though it showed some degree of backsliding: 38% used weed-and-feed. Nonetheless, the study provides evidence that behavior change campaigns can result in sustainable or long-term behavior change.

We would like to highlight one of the “exemplary” studies. In “Education Campaigns to Reduce Stormwater Pollution in Commercial Areas: Do They Work?” (Taylor, A. et al., 2007), the authors discuss the results of a littering reduction campaign in a small commercial area in Melbourne, Australia. To monitor the changes in awareness, attitude, and self-reported behavior of the priority audience associated with littering prevention, the evaluation team used face-to-face survey instruments before, during, and after the campaign. To evaluate changes in people’s actual behavior with respect to littering and littering prevention, the evaluation team observed littering before, during, and after the campaign. They also monitored litter loads entering stormwater by using side entry pit litter baskets at the intervention and control sites to capture pollutants before, during, and after the campaign. The evaluation incorporated self-reported and observed information, used a control group that included an explanation why it was chosen, and monitored the presence of pollutants in the water. In addition, the monitoring process was conducted by staff members who were trained in collecting data. The training, along with a validated rating instrument (the Clean Communities Assessment Tool) and in-depth personal interviews, were helpful for validating the data collection process. They also performed data analysis using

different styles of evaluation to obtain a more comprehensive understanding of key findings and to verify self-reported information.

Conclusions

We conclude by providing several recommendations from observing the trends, common practices, and gaps in existing evaluations. First, we recommend including a control group for assessment purposes. The current evaluation practices suggest that there are significantly more evaluations that collect baseline data than include a control group. This finding is not surprising given that jurisdictions face budget constraints. Nonetheless, this gap jeopardizes the validity of a campaign's results. The use of a control group helps isolate the campaign's effect from other possible explanations for changes in behavior. This is also important for a repeated behavior change campaign that is only based on pre- and post-treatment data from the previous evaluation. Decision makers may decide to repeat a program because of its high success rate in the past, but may not know the success of the campaign was actually due to another, unobserved factor. Likewise, a campaign that may be thought to have been unsuccessful in changing behavior may in fact have suffered from an unobserved common factor.

Second, we recommend addressing issues of selection bias. While half of the collected studies address or acknowledge such issue, we think more evaluations can meet this criterion especially because it is relatively low cost to at least discuss the uptake among the target audience. For example, program evaluators can compare the demographic of program participants with the general residents' demographic of the jurisdiction by reviewing census data. We recognize that social marketing typically encourages practitioners to have a clearly defined and narrowed audience, but an evaluation of the uptake among the target audience is important for scalability or program expansion.