CENTER FOR WATERSHED PROTECTION

Monitoring to Demonstrate Environmental Results: Guidance to Develop Local Stormwater Monitoring Studies Using Six Example Study Designs

Prepared by:

Neely L. Law, Lisa Fraley-McNeal, Karen Cappiella
Center for Watershed Protection

and
Robert Pitt
University of Alabama

CENTER FOR WATERSHED PROTECTION

8390 Main Street • 2nd floor
Ellicott City, MD 21043
Phone 410.461.8323 • Fax 410.461.8324
www.cwp.org www.stormwatercenter.net

AUGUST 2008

Acknowledgements

This Guidance Manual could not have been completed had it not been for support from U.S. EPA grant CP-83282201-0 and Project Officer Bryan Rittenhouse, along with the contributions from many individuals. Over the past twenty years, many researchers have compiled a wealth of insight into the design and analysis of urban stormwater monitoring studies. The need to distill the many years of this experience and expertise into a single manual is challenging at best, and has left some technical details out in order to relay key concepts and elements about stormwater monitoring to local stormwater communities. We hope we have simplified the very technical and scientific terminology and methods so that communities can begin to sort out their monitoring needs. It is without a doubt that the distillation of all the material wouldn't have made much sense without the input of Tom Schueler. Much appreciation is given to the professionals in the field of stormwater monitoring who provided reviews for the study designs. They include: Bill Stack (City of Baltimore), Peter Hughes and staff (USGS Wisconsin), Scott Struck and staff (Tetra Tech), Garry Grabow (North Carolina State University), and Michael Dietz (Utah State University).

Table of Contents

Introduction

1

Introduction to the Manual

This manual presents six monitoring study designs that can be used by Municipal Separate Storm Sewer System (MS4) communities to assess their local stormwater programs. Limited information is currently available in most MS4 communities to determine how well their stormwater programs are functioning by quantifying their stormwater pollutant reductions to protect receiving water quality. The central purpose of this manual is to provide guidance to MS4 communities on developing monitoring studies whose results can help improve their local stormwater programs by getting more pollutant reduction out of the total community stormwater investment.

Monitoring to evaluate the effectiveness of a stormwater program is becoming increasingly necessary. Communities are finding themselves in need of information to meet mandatory permit requirements (e.g. defined pollutant reduction goals) or to justify budgets that support stormwater programs. Monitoring is a requirement for Phase I MS4 communities and can be used to determine progress towards implementation for many of the other MS4 requirements listed in Table 1 for both Phase I and II communities. For example, Phase II communities are required to develop measureable goals to track progress towards implementing each of the six minimum management measures. Most communities' measureable goals are output-based (e.g. number of stormwater treatment practices installed, number of educational brochures distributed), which is useful from a program accounting standpoint but does not allow changes in water quality as a result of these activities to be quantified.

Table 1. NPDES MS4 Stormwater Permit Program Requirements (based on U.S. EPA Title 40 CFR Part 122.21 and Part 122.26 and summarized in Schueler et al. 2004)	
PHASE I COMMUNITIES	PHASE II COMMUNITIES
▪ Stormwater quality monitoring ▪ Mapping of storm drain network ▪ Outfall screening ▪ Removal of illicit discharges ▪ Source identification ▪ Structural and source control measures to reduce pollutants ▪ Erosion/sediment control program ▪ Demonstration of legal authority to control stormwater discharges ▪ Fiscal analysis	Define measureable goals to implement the six minimum management measures to control stormwater to maximum extent practical. The six minimum measures are: ▪ Public education/outreach ▪ Public participation/involvement ▪ Illicit discharge detection ▪ Construction site sediment and erosion control ▪ Post-construction runoff control ▪ Pollution prevention

The reality is that stormwater monitoring to quantify the impact of MS4 stormwater program activities can be very challenging and expensive. Many monitoring studies fail to produce useful data because of poor study design, quality control or data management issues, unforeseen field situations such as vandalism, droughts or floods, or because the study was over-scoped or under-budgeted. This manual is designed to navigate the stormwater manager through these complexities so they can be confident to develop a monitoring study and the results it produces to get the most out of their limited stormwater dollars.

It is recognized that the audience for this guidance manual may not have a lot of experience or expertise in water quality, stormwater monitoring, or statistical methods used to evaluate monitoring data. The manual is primarily written to present the broad concepts and methods behind setting up special monitoring studies, as well as the practical realities of scoping, designing and contracting out studies that can be directly used to improve their local programs in light of the constraints typically encountered. The guidance has attempted to minimize the use of technical jargon and detailed analytical methods. In many cases, information on statistical analysis and experimental design are presented as examples with the recognition that many different approaches can be used and adapted to suit individual situations. Resources are provided for additional information on the more detailed, technical components of the study design. Perhaps most importantly, this document serves as a portal to other stormwater monitoring resources that can be referenced for more information as needed.

Section

2

Basic Steps to Develop a Monitoring Plan

The task of stormwater monitoring and its impact on water resources is a serious consideration for a MS4 community that typically has limited financial and staffing resources and time to commit to such activities. Investing in a local monitoring study requires a process that can break down the larger issue of protecting water quality into manageable components that are addressed on a priority basis. Guidance to develop and implement a monitoring study is outlined in a three-phase approach that is broken down into nine steps. Figure 1 illustrates the specific steps to support each phase. The three phases of this process include: 1) determine local monitoring needs, 2) scope out a specific monitoring study, and 3) implement the monitoring study. While it is beyond the scope of this manual to provide detailed guidance on each of these steps, a series of six monitoring study designs are provided in Section 3 of this manual to help MS4 communities answer specific questions to support the scoping and development of a monitoring study (Phase 2). A general description of the process shown in Figure 1 is provided where elements that are common to all study designs are discussed in more detail. It should be noted that although this guidance is directed towards MS4 communities; other municipalities, state and federal agencies, universities, and watershed organizations that are responsible for implementing stormwater management programs and practices can also use this guidance.

Phase 1: Determine Local Monitoring Needs

Step 1: Define Monitoring Objective
Step 2: Review Existing Studies and Databases

Phase 2: Scope Out a Specific Monitoring Study

Step 3: Select Study Design
Why is the study needed? (e.g. define scope)
What is the basic approach to the study design?

Step 4: Determine Data and Resource Needs
What data should be collected? (e.g. select monitoring parameters)
How much sampling efforts is needed to get reliable results? (e.g. sample size and frequency)
Who will be involved in the study? (e.g. in-house staff, volunteers, consultants)
How much should be budgeted for the monitoring study?

Step 5: Select Study Sites
What factors should be considered when selecting study sites?
What data is needed to characterize site or drainage area conditions?

Step 6: Develop a Monitoring Plan (to include a QA/QC procedures)
What unique sampling techniques and equipment are needed?
What are special data management and quality control considerations?
What monitoring problems can be anticipated?
What are some good monitoring resources to consult?

Phase 3: Implement the Monitoring Study

Step 7: Collect Field Data
Step 8: Perform Laboratory Analyses
Step 9: Evaluate Data and Draw Conclusions

Figure 1. A Three-Phase Approach to Develop and Implement a Monitoring Study

Phase 1: Determine Local Monitoring Needs

The first step to any monitoring study is to define what you want to accomplish by monitoring and how this may be achieved – i.e., the objective. A clear and precise objective is essential to the development of a well thought-out and focused study where the selection of the study design, parameters, and study sites may easily follow. If the objective of the study is loosely stated, there is potential for going off-track and not collecting the data really needed. To make the monitoring study relevant, it is recommended to relate the objectives directly to a management issue. Table 1 presents a list of objectives for the topics addressed by the six monitoring study designs. For many MS4 communities, the establishment of a monitoring program may be directed by programmatic and regulatory requirements or community issues such as:

- Permit requirements,
- Total Maximum Daily Loads (TMDLs) and impaired waters,

- Minimum standards for receiving waters based on designated or beneficial uses, or
- Known community issues such as flooding or beach closures due to bacteria levels.

Table 1. Example monitoring study objectives using the six study designs	
Study Design (SD)	Monitoring Objective
SD1. Quality of stormwater at the outfall	Quantify the nutrient pollutant loadings from residential urban land uses to target measures that reduce loadings from the greatest contributing land uses
SD2. Source area monitoring	Determine the effect of coal tar sealant on Polycyclic Aromatic Hydrocarbon (PAH) concentrations in stormwater runoff from parking lots
SD3. Performance monitoring of individual stormwater treatment practices (STPs)	Provide estimates of runoff reduction and phosphorus pollutant removal effectiveness of bioretention practices
SD4. Implementation and longevity surveys of STPs	Determine the rate of sediment accumulation in pretreatment devices to develop an improved schedule for cleanout practices
SD5. Monitoring public education programs to improve water quality	Determine the runoff reduction achieved by implementing a rainwater capture program on residential land uses
SD6. Cumulative effect of treatment at the catchment scale	Determine the impact of stormwater retrofit practices in two commercial/retail catchments

Before a specific monitoring study is scoped out, it is recommended that a review of existing studies, reports, peer-review journals and databases be completed to first determine if these other information sources may be able to provide the information needed in lieu of conducting a monitoring study. If existing data is not sufficient to address the monitoring objective, it is on to Phase 2 to scope out a specific monitoring study to answer the study objective.

Phase 2: Scope out a Specific Monitoring Study

With the objective defined in Phase 1, the scope for the monitoring study should determine what you want to accomplish. It breaks down the project into manageable tasks and deliverables to more easily and accurately budget the project and eventually develop a monitoring study. Having a well defined and managed scope will help to minimize "scope creep" that results in an incremental expansion of the project and potential cost overruns. All project partners (e.g., funders, contractors) should be in agreement about the terms of the scope.

Next, the data and resource needs to accomplish the scope of work are defined to include the selection of parameters and what other data should be collected. If it is unclear what data needs there are for the study, it is recommended to re-state the objective more clearly as ambiguous data collection needs often lead to insufficient data or not the right data being collected resulting in a failed monitoring study. The level of sampling effort needed to meet the study objectives and how this will be accomplished are the next steps. This is critical as it requires a balancing of expectations and resource availability to include how and by whom the data will be collected (e.g. staff, volunteers, consultants and the type of sampling equipment and supplies needed). Once these elements are addressed and accepted by project staff and partners, a budget can be prepared.

Project scoping and budgeting is an iterative process that strives to find a balance between what can be accomplished in a manageable timeframe and the available resources. An example budget template with basic monitoring program tasks is provided in Table 2. On average, about 15% of the total project budget is allocated to project planning. The budget should allow for more samples to be collected than the estimated sample size due to sampling errors, sampling mishaps and unexpected sampling conditions. For example, if the statistical estimated sample size is 15 runoff samples, you may consider to scope and budget for an extra 20%, or for 20 samples. Additional costs (e.g. 25% of total equipment costs) may be added to account for equipment repair and replacement where high stormflow can dislodge even secured equipment causing irreparable damage. Appendix C provides unit costs for chemical analysis of common parameters to consider. In the end, the project scope and budgeting processes provide a structure that lays out the monitoring study– what needs to be done, how and by whom and the expected outcomes or deliverables.

To help scope and develop a monitoring plan, a presentation of common elements to all of the study designs are discussed below to include:

- Monitoring parameters,
- Data and Sampling Needs,
- Sampling Methods,
- Staffing and labor considerations,
- Safety considerations, and
- Quality assurance quality control (QA/QC) plan.

These elements should be read in tandem with the individual study designs presented in Section 3 of the manual to develop the monitoring study.

Table 2. Example budget items
PLANNING (~15%)
Background research (data acquisition to include previous studies)
Desktop analysis for preliminary site selection and identify data gaps
Field reconnaissance for final site selection
Project scope and sample design
Develop monitoring plan
IMPLEMENTATION (~85%)
Equipment (purchase, installation and maintenance)
Supply costs
Training (staff and/or volunteers)
Sample collection, storage and transfer
Chemical analysis
Data QA/QC
Data analysis and interpretation
Final report

Selection of Monitoring Parameters

The need for a monitoring study may be directed by programmatic or regulatory requirements such that the selection of parameters may be more easily narrowed by:

- Permit requirements,
- Known pollutants of concern in the area,
- Common pollutants associated with land uses,
- Designated or beneficial uses of receiving waters, or
- Community issues such as flooding.

It is essential to select monitoring parameters that will provide the information needed to fulfill the monitoring objective(s) within the available project resources. Collecting data that will not add any 'new information' to the analysis will be superfluous and a waste of much needed resources, given that most monitoring studies have a timeframe of 1 to 3 years with limited budgets. The selection of parameters should be carefully considered as the cost for analyzing pollutants associated with a single stormwater sample is conservatively estimated at $200 for each sample (based on lower end of costs in Appendix C and excludes oil, grease and pesticides). Analysis of only 20 samples for one location can easily cost $4,000 and costs can quickly increase with each parameter added to the study. This does not include costs for staff time to collect and transport these samples.

Table 3 provides a list of common monitoring parameters of interest that local stormwater program managers may consider in each of the six study designs. Additional consideration of the specific type of parameter is needed. For example, the form of a chemical parameter (dissolved, particulate, or total), the particle size distribution of particulate matter or type of analyses (e.g., TSS or SSC), while channel stability can be evaluated through stream cross section measurements, pebble counts or stream gradient measures. Examples of monitoring indicators associated with these parameters are provided in Burton and Pitt (2002) and ASCE and EPA (2002).

Table 3. List of Monitoring Parameters		
Water Quality	**Physical / Hydrologic**	**Biological**
1. Particulate matter 2. Nutrients 3. Pathogens 4. Toxicants	5. Channel stability/enlargement 6. Discharge rate and volume 7. Streamflow (dry weather flow and recharge) 8. Riparian condition	9. Benthic Index of Biotic Integrity (IBI) 10. Fish IBI 11. Aquatic vegetation
Adapted from: Claytor, R. and W. Brown. 1996. *Environmental Indicators to Assess the Effectiveness of Municipal and Industrial Stormwater Control Programs.* U.S EPA. Office of Wastewater Management. Center for Watershed Protection. Ellicott City, MD.		

Data and Sampling Needs

The data needs for a monitoring study are driven by the monitoring study scope and objectives. Major data needs for all study designs include: characterizing site conditions, rainfall data and analysis, and sampling requirements to meet the monitoring objectives. Basic to all six study designs is a list of site information to characterize the drainage characteristics that are used to interpret and analyze data (Box 1). Data needs also include statistical estimates of the sampling effort. Methods to estimate sample size are provided in Appendix A. Data needs are provided for each of the study designs in Section 3.

Box 1. Mandatory Site Information
(as applicable to the study design)

- Contributing drainage area

Precipitation and streamflow records of the receiving waters. Streamflow relevant to on-line STPs.

- Runoff volume for STPs
- Hydrologic characteristic of the drainage area
- Land use, land cover to include % imperviousness (and how they are connected to the drainage system)
- Maps (land use, land cover, facility, drainage area, etc)
- Aerial photographs
- As-built drawing of stormwater control/ drainage infrastructure
- Reported spills and leaks or other activities in the drainage area that may affect flow or pollutant loadings
- Interviews with public works and facility staff / stormwater control maintenance
- Compiled results from prior studies

Precipitation Data and Analysis

Precipitation data is needed to characterize conditions during the course of the monitoring study (e.g. runoff producing rainfall), help interpret data results, and to make informed decisions about what storms to sample.

The installation of a recording precipitation rain gage on-site, or as close to sampling sites as possible, is recommended. Where warranted, a snow gage is recommended for northern/seasonal climates. This may be optional equipment to purchase along with automated sampling equipment. If resources are limited and the source areas are geographically dispersed, manual rain gages commonly sold at local hardware/home improvement stores may be used to supplement recording precipitation gages. Multiple locations throughout the study sites should be used and compared to an automated rain gage. Placement of the gage is critical for accurate results (Box 2).

Box 2. Tips for Proper Rain Gage Placement and Maintenance.

- Locate rain gage away from buildings, trees, utilities that may obstruct precipitation
- Distance from the rain gage to the obstruction should be at least twice the height of the obstruction (building, tree).
- Good exposure, but secured firmly on structure or in the ground
- Location to lessen impact of wind speed on precipitation (e.g. an exposure that dampens the wind speed)
- Top of the funnel is well above the top of the post or similar mount
- Check the gage for clogs or other interferences at least once per week.

From www.srh.noaa.gov/ohx/dad/coop/EQUIPMENT.pdf accessed 9-25-07.

A review of historical precipitation records is recommended to determine what constitutes a runoff event for the study area and secondly to make informed decisions about what storms may be expected to sample. It is important to samples from a range of storm events as pollutant concentrations vary with different storm flows. The use of stratified random sampling of storm events is recommended where the local rainfall conditions are divided into different categories representing different rain depths, rain intensities, seasons, and/or antecedent dry periods (Box 3). A sufficient number of representative events are sampled for each category. The exclusion of specific rain types may result in misleading conclusions. Although very small rainfalls are abundant, they should not be over-emphasized in the sampling strategy as they do not contribute large fractions of the total flows (Burton and Pitt, 2002). Similarly, very large events also do not contribute significant fractions of the total annual flow. Intermediate flows are usually the most significant. Depending on the monitoring objective, it may also be important to characterize the water quality of dry weather flows. This is most relevant for some stormwater treatment practices (STPs) with significant wet storage and, or baseflows (ASCE and EPA, 2002), or where there are known or suspected leaks in aging infrastructure (e.g. water supply loss along transmission lines).

Box 3. Understanding the rainfall distribution in your area

Analysis of historical rainfall information and expected runoff quantities for the different rain classes should be conducted to determine the best approach to capture the range of storm events. A histogram of different rainfall events or cumulative plot (using software such as Excel) as illustrated here is a useful tool to make informed decisions about the storm events to sample (Figure 2). A stratified random approach to storm selection is recommended for monitoring the effectiveness of STPs because their effectiveness is greatly influenced by storm volume and flow rate. With this approach, annual sampling effort is separated by rainfall depth (Burton and Pitt, 2002). For example, using the graphic below, 0.1" rainfall events contributes about 40% of annual runoff events but less than 5% of the total runoff.

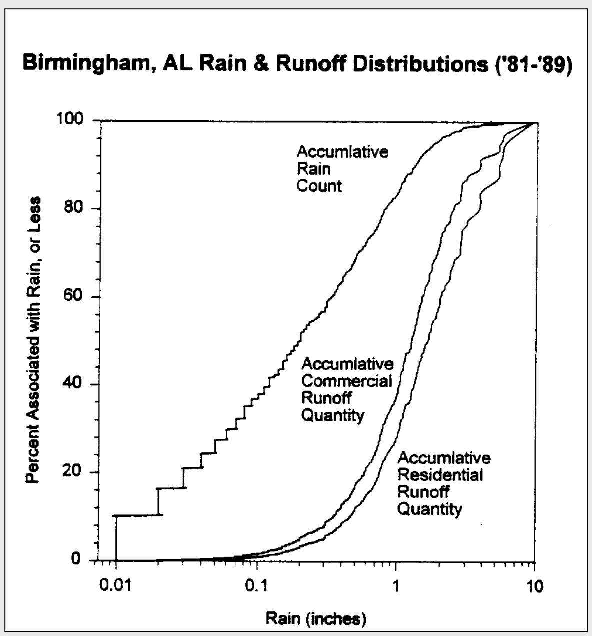


Figure 2. Rainfall and runoff distribution for Birmingham, AL

Level of sampling effort – sample size and frequency

An important aspect of any research is the assurance that the samples collected represent the conditions being tested and the number of samples collected is sufficient to provide statistically relevant conclusions. The ultimate result is to collect useful, meaningful data that reflects the efforts taken to implement a monitoring study. Many times, sample size determination follows traditional "best professional judgments" or is resource driven. However, the results may be unable to fulfill the monitoring objective if insufficient samples are taken or if too many samples collected and resources are wasted. Appendix A provides easy to follow steps on how to estimate sample size needed for a single sampling point (Study Designs 1, 2) or for paired sampling (Study Designs 3, 5, 6) to minimize these possibilities. In general, as more samples are collected, the uncertainty of the overall average concentrations becomes reasonably steady (Burton and Pitt, 2002). 'Power analysis' described in Appendix A is a useful tool to estimate sample size based on the desired outcomes for the study.

An appropriate sampling frequency is determined for each parameter that will yield the desired number of samples in the given timeframe. A reasonable goal for stormwater sampling in humid areas of the country is 10 storms per year where water quality and flow measurements are typically done on a continuous or event-based basis. Most geomorphic, habitat and biological parameters are sampled less frequently and can extend the sampling period. These parameters can be sampled more frequently if desired. For example, if the study focus is limited to macroinvertebrate response, resources may allow for sampling six or more times per year, which will significantly shorten the study timeframe.

Sampling methods: Automated vs. Manual Sampling Methods

The methods of data collection vary across the six study designs to include automated sampling methods (Study Designs 1, 3 and 6), manual sampling (Study Design 2) and visual or behavior survey techniques (Study Designs 4 and 5). These field collection methods are suggested based on the purpose of the study design (e.g. what type of data is needed), the number of sampling sites and their associated sampling costs. In some cases, an automated sampling approach may be used in place of manual approaches, for example in Study Design 2 for source area sampling. A summary of advantages and disadvantages a community may consider for manual versus automated sampling approaches are listed in Table 4.

Automatic water samplers commonly used for stormwater sampling have flexible programming capabilities specifically designed for stormwater sampling. These samplers typically have flow meter and rain gage options so that rainfall data can be logged, along with flow information and sampling history. Newer automatic samplers also have the capability to interface with water quality monitoring probes, which allow specific water quality conditions to trigger sampling. A simpler type of sampler restricted to composite sampling on a time increment basis with little control over sample volumes collected is also available.

Table 4. The Advantages and Disadvantages of Manual and Automatic Sampling (EPA 1982)

Type	Advantages	Disadvantages
Manual	Low capital cost	Probability of increased variability due to sample handling
	Not a composite	Inconsistency in collection
	Point-in-time characterization	High cost of labor
	Compensate for various situations	Repetitious and monotonous task for personnel
	Note unusual conditions	
	No maintenance	
	Can collect extra samples in short time when necessary	
Automatic	Consistent samples	Considerable maintenance for batteries and cleaning; susceptible to plugging by solids
	Probability of decreased variability caused by sample handling	Restricted in size to the general specifications
	Minimal labor requirement for sampling	Inflexibility
	Has capability to collect multiple bottle samples for visual estimate of variability and analysis of individual bottles	Sample contamination potential Subject to damage by vandals

Understanding the capabilities of the sampling equipment and how it will meet the data requirements for the monitoring study is critical. For example, automated sampling equipment may limit the size of particles sampled based on the intake velocity of the sampling tube (most have a minimum intake velocity which limits intake of medium to coarse-sized particles). To address this issue, two alternative sampling options are suggested and include: 1) sampling and analyzing bedload material not sampled by automated samplers, 2) use a cone sample splitter to measure total suspended solids (Pitt *et al.* 2008); or 3) wet sieve the whole water sample described by Selbig *et al.* (2007) to measure the suspended sediment concentration (SSC) (Box 4). Researchers have found cone splitters, in general, work best for sediment concentrations less than 1,000 mg/L and particles less than 250mm as noted in Waschbusch (2003), Horowitz et al. (1997). However, this will be based on the monitoring objectives that determine if it is important to screen for these larger particles or gross pollutants (to include leaves and detritus) and whether or not to include them in sample collection and analyses. This is critical when comparing data to other studies as these gross pollutants can add significantly to pollutant estimates (see Waschbush *et al.* 1999, Selbig and Bannerman, 2007 as examples).

Staffing and Labor Considerations

Field staff play a critical role in any monitoring study as they investigate potential sampling sites, gather site information, collect and transport samples, and maintain sampling equipment. The availability of local program staff to perform these functions, and their training needs, is an important consideration for any MS4 starting up a monitoring study. For example, the daily work schedules of staff can have a significant impact on sampling efforts and expectations for the study implementation should be set accordingly (e.g. staff available to sample storm during off-hours and weekends). An initial estimate of the staff time needed for field crews should be made as part of the scoping and budgeting process. Staffing estimates are provided for each of the six study designs in Section 3.

Options for staffing a monitoring study include:

- The use of existing program staff.
- Transfer employees from different departments to serve on field crews on a temporary basis since some monitoring studies may be seasonal in nature.
- Hire new program staff. Program managers should determine if new hires are anticipated and identify the desired skills (e.g., past monitoring experience).
- Hire interns from local universities and community colleges.
- Solicit the use of volunteers.
- Hire private consultants.

No matter which option is selected, all field personnel must have the appropriate training. For example, 'confined space entry' training may be required under OSHA regulations if sampling will occur in manholes or other confined areas (see Safety Considerations for more information).

Box 4. Measuring Sediment: Total Suspended Sediment (TSS) and Suspended Sediment Concentration (SSC) (CWP 2008)

In many communities, sediment is a primary pollutant of concern and is commonly evaluated using total suspended solids (TSS). The method of analysis for TSS follows those initially developed for wastewater treatment, not natural waters such as stormwater control. Based on the inherent differences in wastewater quality and stormwater quality, such as particle-size distribution, research has demonstrated biases that result from using TSS as a measure to evaluate the pollutant removal efficiency of STPs. Specifically, research has shown that TSS measurement methods used for wastewater analysis applied to the analysis of stormwater can underestimate the amount of sediment in natural waters (e.g., Lenhart 2007, Gray et al. 2000).

Current TSS methods use a subsample of the total sample for analysis. Due to the potential settling of coarse (sand-sized) particles during the sample processing, the subsample is not necessarily representative of the whole sample. "Shake and pour" and pipetting of subsamples frequently misses coarse-sized fractions (e.g., greater than about 250 μm) if they are present in the stormwater (Clark and Pitt, 2008). However, when the subsample is obtained by proper splitting using a cone splitter, then the results can be accurate (Pitt et al 2008).

Suspended Sediment Concentration (SSC) is presented as an alternative and more reliable method to the traditional "shake and pour" method to estimate the amount of sediment in natural waters where the entire sample volume is included in the analyses. This whole water sampling method is suggested to measure sediment strengths in stormwater. (see Gray et al. 2000 and Selbig et al. 2007 for a description of methods).

Use of Volunteers

Using volunteers, such as watershed organizations, homeowners associations and civic groups, to assist with monitoring can greatly reduce costs and is a way to inform the public about local environmental issues and engage community stakeholders in the collection of data that may ultimately affect local decision-making. This can result in a strong sense of stewardship. There are many contributions volunteers can make in a monitoring program to match skills and interest level for participation. Use of volunteers may allow collection of a much larger dataset (i.e., more intensive data collection or more extensive datasets) than the program staff would otherwise not have been able to collect. However, data collected by volunteers to support regulatory-based monitoring efforts may be legally suspect, and regulatory agency officials should be prepared to resample if alerted to unusual conditions that may require court action.

Training is essential when using volunteers and should be incorporated into the program scope and budget. Although there is no fiscal cost for volunteers there is staff time associated with training and providing the necessary materials. The following elements should be considered or addressed when developing a training program for volunteers:

- Ensure volunteers are trained in the proper monitoring protocols.
- Ensure volunteers are trained in the appropriate safety procedures and precautions.
- Provide a description of the volunteer training procedures and 'job description' to clarify the individual roles staff and volunteers have in the field.
- Be clear who the contact person for the monitoring program is and provide this information to the volunteers.
- Be prepared with the needed supplies for volunteer use in the field.

Volunteer efforts may be particularly useful with Study Designs 2, 4, and 5 due to their labor-intensive nature and multiple sampling locations. One key role for volunteers in all of the study designs is to do regular checks of the sampling equipment to determine if maintenance is necessary or vandalism has occurred. The proximity of residents to the sampling sites makes this a natural role for volunteers. Study Designs 1, 3, and 6 require more technical training and experienced personnel due to the automated sampling methods recommended. If volunteers are used for these types of monitoring studies, it is recommended that they be part of a field crew that includes experienced program staff. For all volunteer activities, it may be beneficial to do spot checks of data for QA/QC purposes, or to pair volunteer teams with program staff teams in a particular study area for a day and compare data results to detect any obvious variations in the collection methods. It is also recommended to manage volunteer data as they are received since volunteers may not be available or involved on a regular basis. If there are questions or issues about the data collected, a quick follow-up is best.

A simple on-line search for "volunteer monitoring networks" can reveal numerous organizations in your area that have already trained volunteers or programs to tap into for assistance with monitoring activities (see Section 4 of this manual).

Use of Private Consultants

Another option for staffing the monitoring study when in-house staff capacity or technical expertise is not sufficient is to hire a private consultant. Two major monitoring tasks that are most often contracted out include sample collection and laboratory analysis.

Some considerations when using a consultant to collect monitoring data include:

- Define specific tasks in the scope for the consultant to complete.
- Be sure to clearly identify what information or other support program staff will provide to the consultant (e.g. land cover maps, facility maps) and include this as part of the contract.
- Get cost estimates for each stage of the study and set timelines for deliverables.
- Assign one person to oversee the process if using a combination of consultant and in-house labor.

Look for consultants with the following skills:

- Experience in stormwater sampling and analytical capabilities
- Engineering background,
- Forensic stormwater capability,
- Experience (or at least knowledge of) with monitoring complex STPs would be beneficial,
- Data management & analysis skills, and
- Lab certification and accreditation.

Safety considerations

Stormwater sampling can expose field personnel to hazardous conditions such as high flows, deep pools, soft sediments, etc. Therefore, it is important that field crews have proper safety training and always follow safety precautions. It is recommended that safety training be incorporated as part of the budget scoping process. If sample collection cannot be carried out in a reasonably safe manner, then an alternative approach should be used. Most of the potential hazards can be avoided through the use of automated samplers, careful site selection and selection of sampling times. The following elements are recommended as part of basic field safety training.

- Always sample in pairs and have contact information and identification available to include:
  - emergency contact information from volunteers,
  - identification cards or badges to be worn in the field by field crews, and
  - a 2-way radio or cellular phone and exchange phone numbers with other field personnel at different sites.
- Ensure safety supplies and equipment are available such as:
  - first aid kits in the field,
  - a weather radio to alert field crews to changing conditions, and
  - ensure adequate safety equipment is available.

- Depending on the exposure to open waters, or other hazardous conditions (e.g. steep slopes) the minimum equipment will include:
  - throw rope,
  - inflatable life vests, and
  - nylon covered neoprene waders (that offer some floatation, even when swamped).
- In all cases, only go into the stream if absolutely necessary.
- Try to collect all samples from land, especially if during heavy rains.
- Always enter the water cautiously and be prepared to make an efficient retreat if insecure.
- Never enter a stream where footing is unstable or if the water is too deep (probably more than 2 feet deep) or fast (probably more than 2.5 ft/sec) (Burton and Pitt 2002).
- Wear protective gear to protect from personnel hazards and risk of contamination (gloves, long sleeves and pants, waders if entering the water, and footwear).
- Bring along a good dose of common sense (lifetime supply recommended).

Quality assurance quality control (QA/QC) plan

The QA/QC plan is part of the monitoring study that describes the field and laboratory procedures to be followed in order to limit the errors in sampling or analyzing the data. Although the QA/QC can take a significant amount of up-front planning, if implemented, it can increase the efficiency of the monitoring study by prescribing a set of standard 'rules and procedures' and provide an early detection of problems in the field or the laboratory. Every member of the project team, including staff and field personnel, should have a copy of the QA/QC plan and be familiar with its elements. Appendix G describes the key elements to include in a QA/QC plan. An overview of these elements is provided below.

QA/QC for field sampling

QA/QC procedures for field sampling include 1) determining the storms that are 'eligible for sampling', 2) sample collection and transport, 3) equipment decontamination, 4) field sample containers and labeling, 5) chain-of-custody records and 6) sample receipt. The quality of the sample collected in the field will have a direct impact on the analytical results. It is important that during field collection, enough sample volume should be collected to have material for duplicate and split analyses. A person on the team should be identified as the contact person for the laboratory to track the status of samples, results and address issues that may arise.

QA/QC for Laboratory Analysis

QA/QC procedures developed for laboratory analysis address three major areas: 1) selection of laboratory to conduct analyses, 2) specifications of analytical methods and procedures to ensure the desired results are produced (e.g. use of blanks and split samples) and 3) review of data results to meet data quality objectives. The project manager needs to consult with the laboratory (in-house or contract) to determine what analytical procedures will be used and their detection limits, and to ensure the laboratory has the proper accreditations and technicians have the needed and up-to-date certification for the analyses being conducted.

Phase 3: Implement the Monitoring Study

Implementation of a monitoring study involves three steps: collecting field data, performing laboratory analysis, and evaluating the data to draw conclusions about the findings of the study. The details of each step can be extensive and are highly dependant on the study design and monitoring parameters selected. While it is beyond the scope of this manual to provide guidance on each step, some considerations for implementation are provided within each of the 6 study designs.

A phased approach to monitoring is recommended for each study design. This entails preliminary analysis on an initial set of data to determine if the monitoring set up and equipment is working properly. That is, is the data being collected in the manner that is intended? For example, is the monitoring equipment being dislodged and needs to be better secured, or better concealed due to vandalism? Or, is stormwater runoff entering the sampler that shouldn't (from other source areas) or the intake is frequently blocked by debris? A field notebook records site conditions during sample retrieval and between sample collection times. Modifications of the sampling set-up or equipment may be needed based on these preliminary results.

Data evaluation is the step of the monitoring program where the data are statistically evaluated to determine if the monitoring objectives are met. For example, if the objective of the monitoring program is to determine nutrient levels at outfalls to prevent nuisance algal blooms, the monitoring data generated at an outfall (Study Design 1) may be evaluated to determine if total phosphorus concentrations in the receiving water exceed the U.S. EPA recommended level of 0.1 mg/L, or other locally defined criteria and the statistical significance of the results. Appendix B provides a basic primer to begin statistical analysis of the data generated using the six study designs and provides information on the basic steps to analyze data to determine the 'meaning'. Exploratory, graphical and regression statistical methods are presented.

Section

3

Introduction to the Six Study Designs

The six study designs presented here range in scale from site specific, to watershed wide and, as a whole, address short and long term monitoring needs and a variety of structural and nonstructural stormwater control practices. The full study designs are presented in this Section or may be individually downloaded from the Center website at http://www.cwp.org. The topics include:

Study Design 1: Quality of Stormwater at the Outfall

Study Design 2: Source Area Monitoring

Study Design 3: Performance Monitoring of Individual Stormwater Treatment Practices

Study Design 4: Implementation and Longevity Surveys of Stormwater Treatment Practices

Study Design 5: Monitoring Public Education Programs to Improve Water Quality

Study Design 6: Cumulative Effect of Treatment at the Catchment Scale

In general, the study designs are presented in recommended order based on the status of an MS4's stormwater program. Study Designs 1 and 2 are important for communities that are just starting up their stormwater programs and need to determine the quality of runoff in the community and identify the land use types and source areas that contribute the most pollutants to receiving waters. Study Designs 3, 4 and 5 are good for communities that have established stormwater programs in place and are looking to determine if the STPs they have implemented are a) performing well, b) holding up well over time, and c) if their stormwater education programs are making a difference. Study Design 6 is the most complex and will likely only be implemented by more advanced communities and is the only study that really answers the question of "is what we are doing really making a difference to improve our streams?"

Each of the six study designs is organized by a set of key questions that are designed to help develop a monitoring plan or scope of work for that particular monitoring study. The key questions within each study design will provide the information needed to develop a monitoring plan or scope of work presented in Section 2 of this manual.

- WHAT IS THE STUDY DESIGN? This section describes the overall purpose of the monitoring study design.
- WHY IS THE MONITORING STUDY NEEDED? This section highlights the need for the monitoring study and describes various outputs of the study and how they can address central questions surrounding the effectiveness of stormwater programs.
- WHAT IS THE BASIC APPROACH TO THE STUDY DESIGN? This section provides a basic description of the monitoring study design.
- WHAT FACTORS SHOULD BE CONSIDERED WHEN SELECTING STUDY SITES? This section presents criteria for selecting monitoring sites for the study using desktop and field assessments. The central message is that numerous sites will need to be investigated before a sufficient number of appropriate sites are found.
- WHAT UNIQUE SAMPLING TECHNIQUES AND EQUIPMENT ARE NEEDED? This section describes the unique set of sampling techniques and equipment needed for the study. More commonly encountered sampling requirements are described in Section 2 of this manual.
- WHAT MINIMUM DATA IS NEEDED TO CHARACTERIZE DRAINAGE AREA/SITE CONDITIONS? This section describes the minimum data that should be collected to characterize the sampling site and/or drainage area.
- HOW MUCH SAMPLING EFFORT IS NEEDED TO GET RELIABLE DATA? This section provides guidance on determining the appropriate sample size to generate reliable results for the study design. This section also covers the types of storms to sample and the sampling frequency.
- WHAT ARE SPECIAL DATA MANAGEMENT AND QUALITY CONTROL CONSIDERATIONS? This section describes some data management and quality control issues to consider that are unique to the study design. Section 2 of this manual provides an overview of more general quality control procedures, with more detailed information provided in Appendix G.
- HOW MUCH SHOULD BE BUDGETED FOR THE MONITORING STUDY? This section provides an example budget for the study design based on specific assumptions regarding staffing, monitoring parameters, number of samples and length of study. Unit costs are provided so that the values may be transferred to other studies with varying levels of efforts and tasks.
- WHAT MONITORING PROBLEMS CAN BE ANTICIPATED? Stormwater monitoring is not free of Murphy's Law, where "what can go wrong, will go wrong." This section is provided to hopefully reduce the occurrence of the unexpected by identifying common problems encountered by practitioners in the field.

- WHAT ARE SOME GOOD MONITORING RESOURCES TO CONSULT? This section provides references and web links with more detailed information on the study design and example case studies when available.

Design

1

Quality of Stormwater at the Outfall

What is the Monitoring Study Design?

Stormwater quality monitoring at an outfall is one of the most common types of monitoring conducted by MS4 communities. The purpose of this study design is to develop a statistically reliable dataset to characterize stormwater quality for a particular catchment or land use type. Knowing the types and concentrations of pollutants in the community is an important first step in developing a stormwater program.

Before embarking on development of a local stormwater outfall monitoring study, stormwater managers should consult local, regional or national stormwater quality databases, such as National Stormwater Quality Database (NSQD), that define event mean concentration (EMC) statistics derived from a large population of runoff monitoring samples (Box 1). Depending on the intended use of the data, communities may be able to use local or regional results from the NSQD in lieu of investing in stormwater outfall monitoring. However, outfall monitoring may be needed beyond local stormwater characterization for compliance monitoring to measure the effectiveness of stormwater management programs. Table 1 summarizes Event Mean Concentrations (EMC) from the NSQD from a subset of the more than 20 common stormwater pollutants in the database from residential, commercial, industrial, roadway and open space land uses.

Box 1. The National Stormwater Quality Database

An updated NSQD (version 3) has been completed and is available as a large Excel spreadsheet and described in several summary papers at the website noted below. Nearly 8,500 storm events have been reviewed and summarized in the National Stormwater Quality Database (NSQD). http://unix.eng.ua.edu/~rpitt/Research/ms4/mainms4.shtml. The data are mostly from National Pollutant Discharge Elimination System (NPDES) Phase I communities, who were required to conduct monitoring to characterize stormwater as part of their MS4 NPDES permit applications. Additional data in the NSQD was obtained from the National Urban Runoff Program (EPA 1983), the USGS, and the International BMP Database. The NSQD is an extremely helpful tool to define expected EMCs for a wide range of stormwater pollutants representing different land uses for most parts of the nation (Pitt et al., 2004).

Table 1. Example Pollutant EMCs in Stormwater Runoff from NSQD, vers. 3.0. The value in parentheses is the number of samples.						
	All Data	Residential	Commercial	Industrial	Freeways	Open Space
Median Event Mean Concentrations (mg/L or ppm, except where noted)						
TSS	62 (6,780)	59 (2,167)	55 (843)	73 (594)	53 (360)	10.5 (72)
COD	53 (5,070)	50 (1,473)	63 (640)	59 (474)	64 (439)	21.3 (12)
Fecal Coliform[1]	4,300 (2,154)	4,200 (505)	3,000 (270)	2,850 (317)	2,000 (67)	2,300 (7)
Total P	0.2 (7,425)	0.3 (2,286)	0.2 (920)	0.2 (605)	0.3 (585)	0.0 (77)
Total Cu[2]	15 (5,165)	12 (1,640)	17.9 (753)	19 (536)	17.8 (340)	9.0 (15)
Total Zn[2]	90 (6,184)	70 (1,912)	110 (839)	156.2 (596)	100 (587)	57.0 (16)
Source: http://unix.eng.ua.edu/~rpitt/Research/ms4/mainms4.shtml [1] MPN/100 mL, which represents the most probable number (MPN) of bacteria that would be found in 100 mL of water [2] Cu and Zn values are shown in micrograms per liter						

Why is Local Monitoring at the Outfall Needed?

Stormwater outfall monitoring can be used to:

- Show compliance with NPDES stormwater permits.
- Determine which land uses contribute most significantly to pollutant loadings, which in turn can be used to target pollutant reduction measures and practices to help meet Total Maximum Daily Loads (TMDLs) or other water quality goals.
- Provide information to select and design stormwater treatment practices (STPs) to achieve greater reduction of pollutants of concern beyond the median removal rates (refer to CWP 2007). The NSQD data can be used for preliminary designs and cost estimates, but it is also important to invest a modest confirm local discharge conditions before expensive controls are designed and implemented.
- Provide locally-derived runoff and water quality data from representative land use catchments needed to calibrate parameters for use in simulation models to predict hydrology and water quality conditions across a community.
- Determine the effect of stormwater treatment on runoff characteristics at the outfall as part of a paired watershed (see Study Design 6 for more on paired watershed studies).
- Define a runoff coefficient for a specialized land use that cannot be readily characterized by an existing coefficient.
- Accompany source area monitoring to estimate the proportion of pollutant load attributed to specific source areas as compared to the total catchment load (see Study Design 2 for more on Source Area Monitoring).

Communities can use existing data from the NSQD to meet many of these management objectives. However, the database has some limitations, which may necessitate development of a stormwater outfall monitoring program for certain situations that are not well-represented in the NSQD. These include:

- Regions with limited NSQD data coverage: such as the arid SW and northern mountain states where few Phase 1 stormwater permits have been issued.

- Land uses and conditions that are not well represented in the NSQD: such as open space (parks, golf courses), construction sites, and snowmelt.

- Water quality parameters that are poorly represented in the NSQD: hydrocarbons, bacteria, organic carbon, pesticides, and PAHs.

What is the Basic Approach?

With an outfall monitoring study, stormwater quality and flow are monitored below the outfalls of small catchments with homogenous land use. Sampling is conducted throughout the entire storm hydrograph over many storm events. Runoff quality for the individual catchment or land use type is ultimately characterized using an EMC and load, when combined with flow data, for the pollutant(s) of concern. Statistical analyses may reveal important factors affecting the observed concentrations for a monitoring site, such as effects by season, rain depth or intensity, or an intervening dry period.

The basic approach is to sample runoff from a pre-determined number of outfalls that are representative of different land use types in the community during storm events in order to characterize runoff from each land use type. Outfall monitoring may also focus on a smaller paired-sampling approach to compare the effect of treatment or restoration between two similar land use catchments. Weirs or other structural devices may be needed to concentrate the flow from an outfall (Figure 1). Outfall samples should be taken where the water is cascading onto the sample intake. Placement of the sample intake on the bottom of a pipe easily clogs with debris and often over-represents the bedload component of stormwater. However, placement of the intake above the bottom of the pipe may lead to an under-representation of larger stormwater particulates. The best location is therefore where the water is actually falling onto the sampler intake and is therefore completely mixed. Preliminary analysis from a phased monitoring approach can help to get an idea of local variability in stormwater and as a result, adjust the sample size if needed.

What Factors Should be Considered when Selecting Test Sites?

The goal is to find representative, homogenous land use catchments. Multiple sites for each type of land use catchment may be needed and will ultimately depend on the study objectives. This starts with a survey of the types of land uses within the study area that are screened as potential sites using desktop analysis techniques, aided by GIS if available. An initial list of land use areas to be considered for monitoring should be based on available land use maps, but additional information through the use of overlay techniques provide more information about the stormwater quality (e.g. age of development that can be a surrogate or lead to information about the degree of connected impervious cover, presence/absence of stormwater practices and type.) For example, Wright et al. (2004) and Bochis et al. (2008) provide field methods to characterize land cover and management practices within specific land uses. Consult Burton and Pitt (2002) for methods for cost-effective site selection methods based on expected discharges and the mapped distribution of the different land uses in a community.

Figure 1. Example outfall monitoring with V-notch weir and flume set-up. (photo courtesy City of Baltimore, MD)

The following screening factors may be used to guide selection of initial test sites using desktop analysis:

- Catchment size is between 25 and 200 acres (although it can be larger given study design objectives or local development patterns),
- At least 80% (and ideally as much as 100%) of the catchment is comprised of the land use type of interest,
- Outfall size is between 24 to 48 inches, and
- Outfall is adjacent to public land for easy access and to house monitoring equipment.

Additional screening for final selection of monitoring sites will require site visits to evaluate specific site conditions that may affect sampling. During site visits, stream conditions above and below outfall (up to 1 mile minimum) should be assessed and recorded. This includes channel morphology, flow, habitat, riparian zone, sediment type, organic matter, oil sheens and odors. Observations may also record biological communities and other parameters if this serves the study objectives. Additional considerations for selecting study sites are listed below.

- Site should be stable with good access, availability of power (if needed), and space to accommodate monitoring equipment.
- Select sites that are in close proximity to facilitate the efficiency of sample collection and equipment set-up for storms.
- Sites should have limited public access to minimize potential for vandalism.
- Sufficient flow that is well mixed (not stagnant).

- A stable cross-section with well-mixed flow is needed to get flow rating to calibrate the flow sensor equipment (i.e., estimate stormflow based on stream level/stage). The flow sensor needs to be located upgradient from the sampler intake in a stable flow section.
- Consider whether the site is appropriate for the equipment selected. For example, automated samplers are restricted to a vertical height of about 7 meters from the water surface to the sampler pump. If the sampler height is greater than this critical height a submerged pump can be used to solve this problem, at additional cost.
- Stream daylight situation right below the outfall.
- Avoid wide shallow and fast flowing streams, as they may be difficult to sample adequately.
- Avoid catchments with active construction (unless construction sites are a part of the study design).
- Avoid sites with steep slopes or adjacent traffic that makes access hazardous for crews.
- Avoid sites with dry weather flow in pipes (check for illicit discharges using Brown et al., 2004). Even clean dry weather flows in pipe will interfere with sampling. However, if the monitoring study design objective is to characterize illicit discharges, sampling at the outfall should include these sites in the study.
- Avoid sites where there is evidence of trash/debris accumulation near the stations. This may indicate access and use of the site by the public and may interfere with the sampling equipment.
- Avoid depositional zones, such as river bends and mouths, pools, and impoundment structures and those influenced by tidal fluctuations. These areas may be sampled for sediment contamination and toxicity as a special component to the monitoring program.

What Unique Sampling Techniques and Equipment Are Needed?

Automated sampling is recommended but may not always be necessary depending on the purpose of the monitoring study. Two situations that would require grab samples include: 1) if the focus on the study is to characterize first flush samples, or 2) if the pollutants of concern require grab sampling, such as bacteria, oil and grease, or volatile organics. If automated samplers are used, the equipment should be installed so that it is secure and stable to prevent damage during high flows. Other required monitoring equipment includes flow sensors or other flow measurement devices and a tipping bucket rain gage. The tipping bucket may be connected to the automated sampler to initiate sample collection.

The preferred sampling method is to obtain flow-weighted composite samples. Automated samplers may also be programmed to take multiple, discrete samples throughout the storm, but this is more costly and results in much more data to manage.

One situation where this might be desirable is when the objective of the study is to look at the effect of different flow intensities on pollutant concentration. To determine the variability of stormwater quality during an event, one method is to take discrete samples for a key parameter that is well correlated with other pollutants. One such parameter is turbidity because many pollutants are associated with particulates. After the key parameter has been measured from the discrete samples, the samples can then be manually composited into one sample, from which the more complete list of constituents of concern can be measured. By correlating turbidity with the additional constituents measured, the variability of the constituents throughout the storm event can be estimated. Another efficient method to measure varying stormwater quality during an event is to use a real-time sonde that can measure turbidity, DO, temperature, pH, and conductivity every few minutes during an event. Figure 2 shows high-resolution turbidity plots of the influent (baffle) and effluent (sump) for a stormwater control device during a 6 hour nighttime rain event.

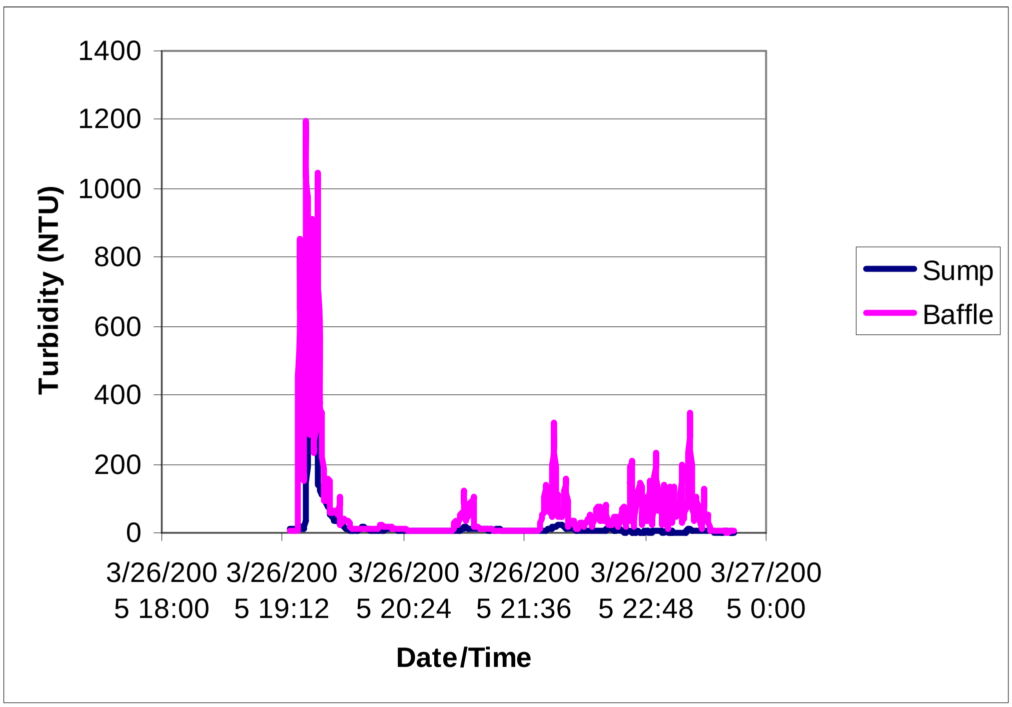


Figure 2. Example High-Resolution Sonde Measurements Comparing Influent and Effluent Turbidity Values during Field Evaluation Tests of the UpFlow Filter using a YSI 6000 series sonde (Pitt and Khambhammettu 2006).

Other special sampling considerations include:

- Ensure the samples are well-mixed through proper placement of the sampling equipment and the samples are representative of the storm event by collecting samples throughout the storm event. Experience is the best trainer where field personnel will need to estimate the size of the storm event to set up the equipment depending on the type of sample collected (e.g. preset time intervals, or volumes).

- Use sampling techniques to reduce the bias of stormwater sediment concentrations and loads by collecting a representative sample of all particle size classes in storm flow.

What Minimum Data is Needed to Characterize Drainage Area/Site Conditions?

The proportion and distribution of land cover types across a catchment can vary greatly among catchments, even when they have similar land uses. Such variations can have a significant impact on stormwater quality at the outfall, particularly for smaller catchments. Therefore, the need for accurate estimates of land cover is paramount to select representative catchments for a stormwater outfall monitoring study. The use of high resolution (e.g. 1m) aerial photographs digital imagery and topographic maps can supplement site visits to discern these variations. Example imagery for two ultra-urban catchments as part of a street sweeping study in Baltimore, MD and an accompanying table of land cover characteristics illustrates the level of detail for this study design (Figure 3, Table 0).

In addition to land cover distribution, mitigation strategies to reduce pollutant loadings within the catchment are important features to note. These include the location and type of STPs, use of environmental site design practices, and method(s) of stormwater conveyance (especially curbs and gutters vs. grass swales). These factors may affect stormwater quality and quantity and should be noted to help interpret variations in concentrations from similar land uses. A summary of basic information needed to characterize the drainage area includes:

- Catchment size and boundaries,
- Predominate land use type and distribution across catchment,
- Land cover distribution across catchment (e.g., percent impervious cover, turf, forest, exposed soil),
- Development characteristics (e.g., age, traditional versus environmental site design),
- Type of conveyance system (e.g., open or closed channel) leading to sampling point,
- Presence and type of STPs,
- Rainfall data, and
- Sampling season to account for seasonal impacts on stormwater quality. Snowmelt should be sampled in northern areas as an additional component of the stormwater annual mass discharges.

Catchment F Monitoring Station at Lanvale St., Baltimore, MD

Catchment O Monitoring Station at Baltimore St, Baltimore, MD

Table 2. Catchment characteristics (CWP 2006)

Characteristics	Catchment F	Catchment O
Total Area (Acres)	38.43	38.70
Impervious Cover (%)	67.8%	76.6%
Pervious Cover (%)	32.2%	23.4%
Streets and Alleys – acres (% of catchment)	10.17 (26.5%)	10.06 (25.6%)
Paved Right of Way [1] acres (% of catchment)	5.79 (15.1%)	5.72 (14.8%)
Rooftop Cover – acres (% of catchment)	9.56 (24.9%)	12.64 (32.2%)
Other Impervious Cover [2] (% of catchment)	0.53 (1.4%)	1.24 (3.2%)
Street and Alley Length (miles)	3.57	3.60
Current Curb Miles Swept Per week [3]	7.69	4.43
Proposed Curb Miles Swept Each week	4.15	11.14
Sweeping Treatment	Restricted	Expanded
Number of catch-basins [4]	92	74

Notes:
1 Sidewalks from edge of street to rooftop
2 Parking lots and driveways
3 Curb miles on each side of street (e.g., 2 times street length
4 Estimated from KCI (2004) SWMM Block modeling
Sources: CWP 2005, KCI, Inc 2004, Stack, pers. comm

Figure 3. Monitoring sites for two ultra urban catchments in Baltimore, MD (★ is the monitoring location)

How Much Sampling Effort Is Needed to Get Reliable Data?

Statistical analysis of existing or reported data (e.g. NSQD) may be used to provide initial guidance to 'ballpark' the number of samples needed and to establish a realistic monitoring budget. Adopting a phased approach to sampling as described previously can help determine if this initial estimate should be adjusted. Appendix A provides example calculations to estimate the number of samples needed for comparisons between different sites or times.

At least two years of monitoring is needed in most regions of the country to get about 20 -30 good "keeper" samples. Longer sampling periods may be needed in arid and semi-arid regions. This is due to the relatively limited time window during which staff are available to collect and process samples (e.g., generally only 40 hours a week). Experience also finds that there is potential to discard about half of the storm events sampled due to unexpected stormwater sampling conditions and sampling errors (e.g., poor flow data, not enough flow, not enough sample bottles, rejects due to quality control, or damage to sampling equipment). The most reliable research monitoring efforts can sample 90% of the annual flows, but that requires significant effort and costs.

A range of different storm events to sample should be identified so as not bias dataset with frequent smaller storms and to avoid the first flush. An analysis of the frequency of storm events from long-term databases or other local datasets takes some guesswork out of this estimation and provides a realistic idea about the number of storm events, on average for the area and the type of storms encountered. However, extremely wet or dry years are likely.

Based on the objective of the outfall monitoring study, if paired samples are needed, Burton and Pitt (2002) suggest as general guidance to collect a minimum of 12 paired samples in order to obtain a reasonable estimate of the variation between catchments and to provide reliable estimates for analyses.

What Special Data Management and Quality Control Issues Can Be Expected?

The goal of this study design is to collect rainfall, water quality and flow data for a range of storm events at the outfall. It is recommended to follow the basic elements of a QA/QC plan for the study design provided in Section 2 and Appendix G of this manual. In addition, the following are specific considerations for monitoring at the outfall.

- Decide how to address gross pollutants to include coarser size particles in addition to organic debris and trash.
- Placement of the automatic sampler intake may bias the collection of coarse size particles and miss suspended particles during higher flow.
- Common errors relate to flow measurements when debris can interfere with the measurements. Post event calibration checks of flow monitoring equipment can add confidence to the results. For example, field staff can compare a chalk line before the event and determine if the maximum stage record corresponds to where the chalk line was washed away. Alternatively, staff may check the baseflow stage recorded by the flow meter and compare it to the field staff gage that is installed for this purpose.

- Careful record keeping and field notes are needed to minimize sampling error and can help to explain sample variability when reviewing the data.

How Much Should be Budgeted for the Monitoring Study?

Table 3 outlines the expected cost for each phase of a typical outfall monitoring study. The table provides some rule of thumb multipliers and unit costs that can be used to adapt these costs for an individual monitoring scenario.

The illustrated budget estimates a total cost of $61,700 for monitoring 3 small residential outfalls draining areas between 150-175 acres within a 10-square mile subwatershed. The objective of the monitoring study is to characterize stormwater quality in medium-density residential neighborhoods (e.g. townhomes) that predate existing stormwater management design guidelines to identify the need for improved STPs that address water quality. The parameters of interest include a standard parameter set of TSS, TN, TKN, nitrite, nitrate and TP. If additional land use types are included in a monitoring study, additional resources would be needed to identify sample sites, and collect and analyze samples.

Table 3. Example budget

		Unit Cost	Total Cost
10 square mile subwatershed			
Goal: 3 monitoring sites, 10 "keeper" storm samples, single land use			
PLANNING (11%)			
Desktop Analysis			
Review existing studies/databases	3 days	$50/hr	$1,200
GIS analysis (incl data compilation)	5 days	$50/hr	$2,000
Study Design			
Site visits (e.g. 30 sites, 10 sites/day)	6 staff days	$50/hr	$2,400
Sampling design	2 day	$50/hr	$800
Training	6 staff days	$50/hr	$2,400
IMPLEMENTATION (89%)			
Monitoring Equipment & Installation			
Automated			
ISCO (water sampler and flow meter)		$10,000	$30,000
Supplies[1]			$1,000
Sample collection[2] (20 samples, 3 sites)	2 staff, 2hrs per site	$50	$12,000
Analysis			
Sample analysis [3](standard parameter set)		$130	$3,900
Data QA/QC	5 days	$50/hr	$2,000
Data analysis	10 staff days		$4,000
			$61,700

[1] $1000 supplies plus 25% replacement cost for equipment
[2]Site visit includes equipment maintenance, travel to site and sample delivery
[3]Due to sampling errors, 10 of the 20 samples collected are analyzed

What Monitoring Problems Can Be Anticipated?

- Equipment repair or replacement due to vandalism, damage from high flow events, or malfunction.
- Debris accumulated behind the equipment can create a surcharge effect, or material downstream that blocks flow can create a backwater effect, both leading to erroneous flow measurements.
- Insufficient samples collected due to lack of runoff-producing storm events, sampling errors, or storms occurring during "off" hours.
- Samples not analyzed because criteria not met (e.g., sample contamination in field or lab, insufficient sample volume, holding time not met).
- Errors or malfunctions during lab analysis.
- Difficulty finding catchments that are representative of a particular land use type (e.g., 80 to 100% of catchment).
- Insufficient data to accurately characterize land use and land cover in catchments.

What are Some Good Monitoring Resources to Consult?

Brown, E. D. Caraco and R. Pitt. 2004. *Illicit Discharge Detection and Elimination: A Guidance Manual for Program Development and Technical Assessments.* Center for Watershed Protection, Ellicott City, MD. 176 pp.

Burton, A., and R. Pitt. 2000. *Stormwater Effects Handbook: a Toolbox for Watershed Managers, Scientists, and Engineers.* Lewis Publishers. New York, NY. http://unix.eng.ua.edu/~rpitt/Publications/Publications.shtml

Center for Watershed Protection (CWP). 2007. *National Pollutant Removal Performance Database. Version 3*. Center for Watershed Protection, Ellicott City, MD 21043.

Digital aerial photographs can found at TerraServer USA http://terraservice.net/

EPA Monitoring and Assessing Water Quality website http://www.epa.gov/owow/monitoring/

National Stormwater Quality Database (vers 3) http://rpitt.eng.ua.edu/Research/ms4/mainms4.shtml

Design

2

Source Area Monitoring

What is the Monitoring Protocol?

Source area monitoring involves the collection of samples to determine representative pollutant concentrations from specific landscape features in a community. This data can help MS4 communities to understand the sources and strengths of their pollutants of concern and can be used to identify the most effective type and placement of stormwater treatment practices (STPs) to reduce impacts on the watershed. A source area is defined as an area with homogenous land cover that is no more than several hundred or thousand square feet in area and transmits runoff as sheetflow or shallow concentrated flow to the stormwater drainage system. Source areas typically include roofs, streets, parking lots and other paved areas, lawns and other pervious areas.

The nature and distribution of pollutants vary among source area types (Table 1), while the contribution from different source areas may vary throughout a storm event and with storm intensity. For example, the contribution of impervious surfaces such as roads, rooftops and streets may provide the initial loadings, while runoff from landscaped areas or other areas with compacted soils may become more significant as rainfall depth increases (Burton and Pitt 2002).

Table 1. Source areas and their associated geometric mean pollutants concentrations (data from Steuer et al. 1997).

Source area	TP (mg/L)	Total Solids (mg/L)	Fecal coli (C/100mL)	Zn (μg/L)	Cd (μg/L)	Cu (μg/L)	PAH
Lawns	2.33[1]	395	4,700	--	--	--	--
Residential Rooftops	0.08[1]	81	2,200	318	0.7	10	0.61
Commercial Rooftops	0.09	115	30	348	0.9	23	2.06
Commercial Parking Lots	0.21[1]	240	4,200	178	0.9	25	75.58[2]
Streets	0.14-0.29[1]	300-498	280-2,400	78-256	0.6 – 1.0	15-31	1.72-15.18

[1] Concentrations of TP reported in Washbusch et al. (1999) report concentrations in lawns (0.79 – 1.66), streets (0.18-0.40 mg/L) and parking lots (0.10 mg/L)

[2] PAH dissolved concentrations in parking lots with a coal-tar vs asphalt sealant differ by an order of magnitude (refer to Mahler et al. 2004)

Why is Source Area Monitoring Needed?

Source area monitoring studies have demonstrated that different urban source areas contribute disproportionate amounts of a specific type of pollutant to receiving waters. Monitoring source areas provides a means to target source control efforts that prevent pollutants from entering stormwater runoff. This technique may be used to:

- Define which urban source areas generate high concentrations of the pollutant of concern and target pollution prevention education towards the behaviors generating pollutants in these areas.
- Evaluate whether pollutant concentrations from a source area vary based on the specific characteristics of the source area (e.g., parking lot design standards, lawn soil type).
- Compare the results of source area monitoring studies to prior studies to evaluate progress in source control efforts, and recommend changes to the local program if needed.
- Provide input data to models that estimate pollutant mass balance in a watershed.
- Use in conjunction with other monitoring efforts to evaluate the effectiveness of non-structural stormwater management practices (e.g., education programs, street sweeping). See Study Design 5 as an example.

Some specific examples of source area monitoring study applications are described in Table 2.

Table 2. Example Source Area Monitoring Studies

Study Objective	References
Evaluate nutrient runoff from lawns in different neighborhoods or under different levels of stewardship or management practices	*Bannerman et al.* 1993, Steuer 1997, Garn 2002
Determine the pollutant strength of various stormwater hotspots.	*Scanlin and Feng* 1997, Mahler et al. 2003
Document the pollutant concentrations of hotspots such as dumpster juice, car washing, etc.	*Dengler and Brasino*. 2007
Evaluate the effectiveness of street sweeping or storm drain cleanout programs by monitoring street runoff quality under different management practices	*Pitt and Bissonette* 1984, *Waschbusch et al.* 1999, *Selbig and Bannerman 20*07, CWP 2006.
Determine the pollutant loading from roofs	*Bannerman et al.* 1993, *Bucheli et al.* 1998, *Polkwoska et al.* 2002, *Chang et al.* 2004

What is the Basic Approach?

In its most basic form, this study design has the goal of determining the types and concentrations of pollutants in runoff from different source areas. To accomplish this, water quality samples are taken throughout the storm event from pre-determined source areas that are representative of the study area to characterize pollutants in runoff. Data can be used to determine if the sources are distributed throughout the watershed or if the sources of pollutants are localized to 'hotspots'. More detailed studies may be done to evaluate whether runoff characteristics vary based on the characteristics of a specific source area (e.g., parking lot surface material, lawn fertilization practices, age of rooftops).

Manual sampling techniques are commonly used for source area monitoring where samples are collected before runoff enters the storm drainage system. Automated sampling may also be feasible given special sampling considerations (e.g., Roa-Espinosa and Bannerman 1994). It is typically inefficient to use the standard automated sampling set-ups (as described in Study Design 1) given the small scale of the sampling areas and the limited locations where the samplers can be placed ensuring that they would only receive runoff from the isolated area of concern. Further, source area sampling involves multiple sampling sites, which may make it cost prohibitive to purchase automated sampling equipment or inconvenient to rotate equipment between sampling stations. Many samples and replicates are generally required to fully characterize runoff from source areas due to sampling errors, and storm and site variability.

Once an MS4 has identified their pollutants of concern (e.g. based on permit requirements, impaired waters, TMDLs, water quality monitoring results), they can use Table 1 to target specific source areas for monitoring and begin to screen the community for source area sites.

What Factors should be Considered when Selecting Test Sites?

Key to monitoring source areas is the selection of representative source areas. A representative source area has characteristics that are typical, or common to the overall study area or community. For example, if the majority of parking lots in the community are asphalt lots designed with curb and gutter with no landscaped islands, representative parking lots should also have these characteristics.

If the study objective is to determine whether runoff characteristics vary based on the characteristics of a specific source area, sampling sites should be stratified by a specific characteristic (e.g. roof type, lawn fertilizer application rate, street traffic volume). In this case, a control site is also recommended. As an example, a monitoring program developed to determine the effectiveness of street sweeping in parking lots would include parking lot sites stratified by pre-determined levels of treatment (e.g., various street sweeping or type of sweeper) and a control parking lot with no street sweeping.

If the study objective is to characterize the strength of a specific type of pollutant or a specific source such as coal tar as a source of PAHs for example, targeted sampling of specific areas is recommended, such as newly paved or resealed asphalt parking lots. Replicate sites for each source area are needed (between 3 to5).

Additional considerations for selecting study sites include:

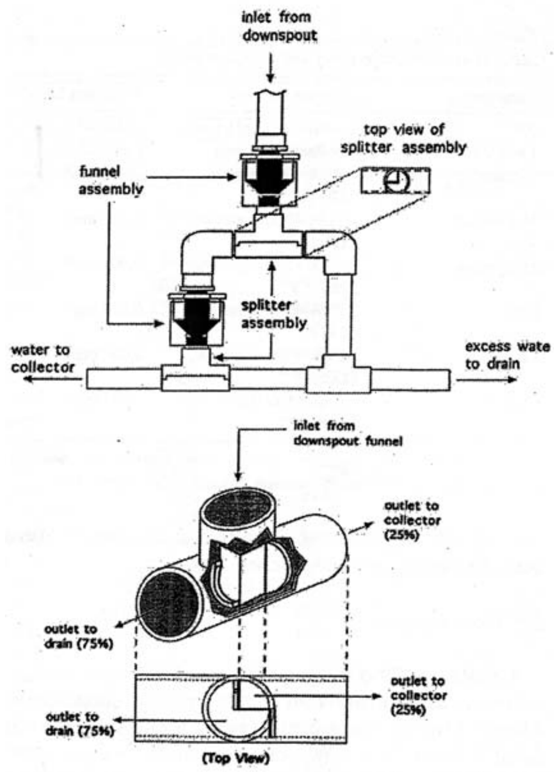
- Realize that only 1 in 5 locations investigated may be potentially suitable sampling sites. You will likely need to investigate a large number of sites to identify a sufficient number of samples sites that meet the criteria for the study.
- Locate test sites to collect runoff only from the source area of interest. The samples collected should "isolate", or contain only stormwater runoff from the specific source area. For example, street samplers should avoid gutter flow that may have originated from other source areas such as driveways, or lawn runoff that may include roof runoff from a disconnected roof leader.
- Cluster sample sites to reduce climatic variability and to make sample retrieval more efficient (e.g., there is likely only a single crew available to retrieve all samples within short period of time, 3-4 hrs, following a storm).
- Sites should have a secure location to install equipment to prevent vandalism or destruction from large storm events.
- Sites should be free of hazards that may interfere with sample collection, retrieval and site maintenance (e.g. parked cars interfere with runoff collection from street samplers).
- Safe access to site is also essential.

The project management or lead for the monitoring study is encouraged to establish a relationship with the property owners prior to finalizing the sampling sites. It may be possible to solicit their involvement in the study to check on the equipment and report any problems on a regular basis, or at least willing to have crews on-site at all hours. The involvement of the property owners does not shift the onus of equipment maintenance and repair on the property owner, but may be helpful since staff or other volunteers who are unable to attend each site every day. Weekly site visits are recommended by study personnel, or more frequently on an as-needed basis.

What Unique Sampling Techniques and Equipment Are Needed?

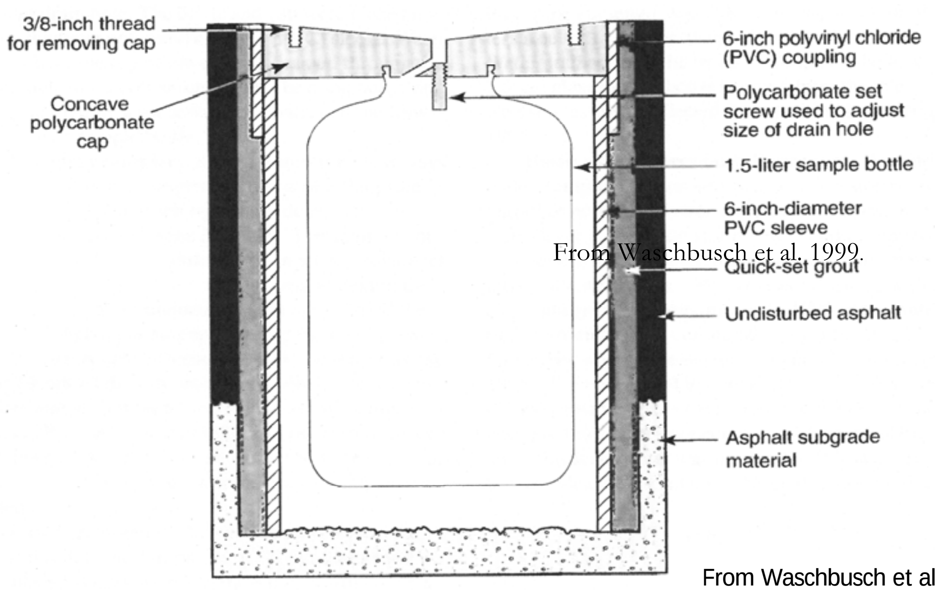
Source area sampling techniques range from simple grab samples that only require a sample bottle to more complex equipment that funnels sheetflow into a container, or extracted using a hand pump. The design of the sampler is based on the monitoring objective such that appropriate collection devices are used. For example, the appropriate sieve or screen size should be included on a sampling device if the objective is to capture gross pollutants such as trash, litter, and large sediment particles. Examples of different source area equipment are illustrated in Figure 1 along with a brief description of how they work in Table 3. For shallow runoff, a manual or more expensive battery-powered peristaltic pump may be used. For example, Mahler *et al.* (2004) used a peristaltic pump to collect runoff in a parking lot that was "pooled" downslope using boards to which weather stripping was applied, or with urethane spill berms. The design and construction of the sampling equipment may need to be modified once in place, given the potentially unique features of a source area sampling site. Close attention to how the sampling equipment works in practice is needed to determine if any modifications are required.

a.)



From Chang *et al.* 2004

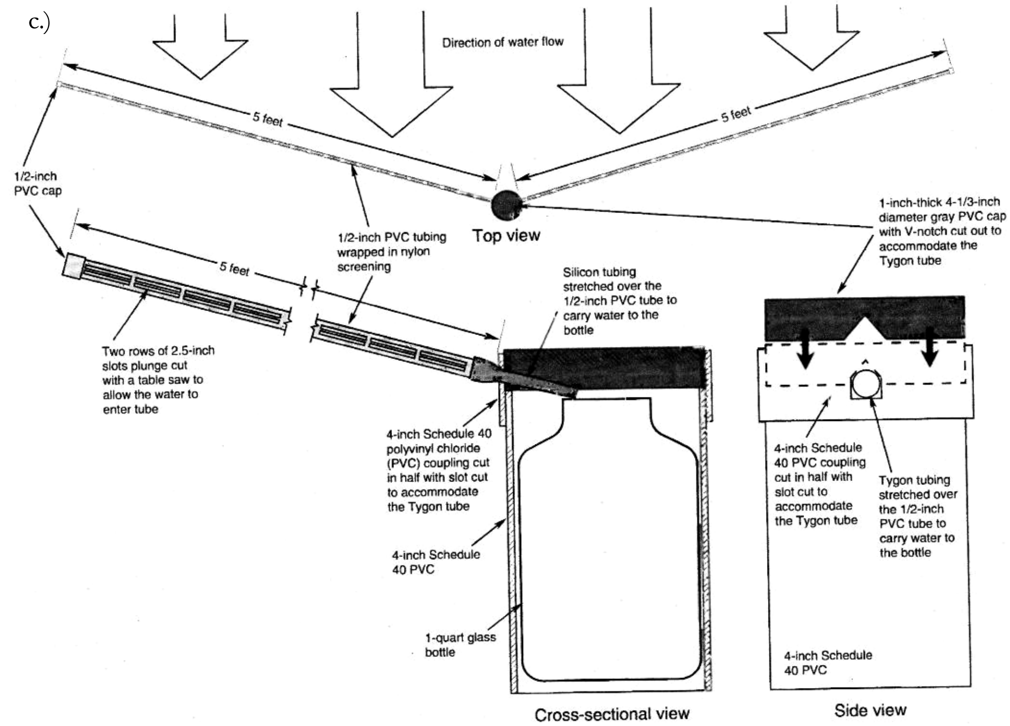
b)



From Waschbusch et al. 1999.

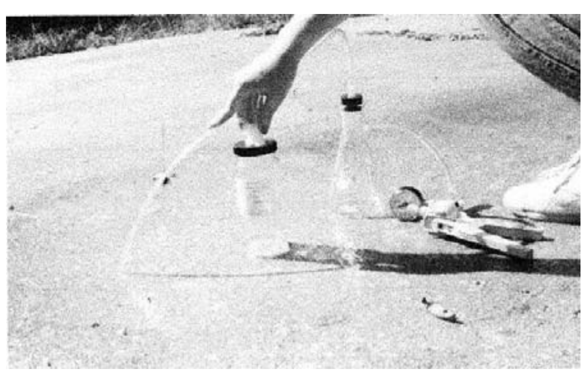
From Waschbusch et al. 1999.

c.)



From Waschbusch et al. 1999.

d)



From Burton and Pitt 2002.

Figure 1. Example source area monitoring equipment for a) roofs, c) streets and c) lawns, d) hand pump.

Table 3. Description of source area sampling methods as described by *Steuer et al.* 1997, *Waschbusch et al.* 1999, *Garn* 2002, *Burton and Pitt* 2002.	
Lawn samplers	Runoff is collected through two, 5-ft pieces of ½ inch diameter PVC pipe place flush with the surface of the ground on a sloped surface, with an angle of about 150 degrees between the two pipes (Figure 2c). Runoff entered the pipes through two slits cut the entire length of pipe. Each pipe was wrapped with fiberglass screen to prevent insects and large debris from entering. The pipes were secured with wooden clothespins and nylon rope. A 1-quart sample bottle collected the runoff from the pipes. The cap had a notch to accommodate silicon tubing, which ran from the end of the PVC collectors to the sample bottle.
Street samplers	Street samplers are grouted into the street approximately 5 ft from the curb (Figure 2b). The sample bottle is covered with a 6-in concave polycarbonate cap, set flush with the street surface, with a center drain hole. The bottle and cap were placed into a 6-in diameter PVC sleeve. Water flowed over the top of the cap and drained through the center hole into a collection bottle. The drain hole could be constricted by a set screw that controlled the flow.
Roof samplers	A portion of the rooftop runoff is diverted to a sample bottle using a ¼ ″ diameter vinyl tube attached to the inside of the downspout using plastic wire ties. Each tube was inserted into a 1.5 L glass sample bottle that was placed in a protective 10-inch diameter PVC sleeve. Overfilling of the bottles may be a problem and can be addressed by controlling the volume of water entering using a polycarbonate cap similar to the street samplers.
Parking Lot Samplers (No image provided)	Runoff enters a storm sewer inlet grate where a small portion of inlet flow is diverted to a sample bottle using a 6-in trough made of ½ inch diameter PVC pipe cut length-wise and held in place with stainless steel hose clamps attached to the inlet grating. Water drains from the trough through a tube to a 2.5 gallon sample bottle hanging from the inlet grate.

General guidance on source area sampling techniques includes:

- Advance planning is recommended to ensure the needed sampling equipment, supplies and field forms/log book are ready to go when a storm event occurs.
- All grab samples should be collected when at least 24 hours of no measureable precipitation has occurred prior to the storm event.
- Maintain contributing drainage areas free of debris or other obstructions that may interfere with collection of representative source area samples. Sampling techniques should have materials such as screens to prevent debris, insects and trash from entering the sample collection device. Take care not to use too small screening that prevents particulate collection. If the monitoring objective is to characterize trash, litter, organic material in the source area, then this material should be included in sampling efforts.
- Avoid collecting first flush samples only to determine the concentration or load. Samples should be collected through the entire event, including first flush. Pollutant concentrations vary during a storm event and it is important to take multiple samples throughout the storm.

- Ensure sample bottles do not over-fill Event samples require personnel to be present throughout the event to replace near full source area sampling bottle with new, pre-cleaned empty bottles.
- For shallow sheetflow use a peristaltic hand operated vacuum pump to create a small vacuum in the sample bottle. Gently draw the sample directly into the container through a Teflon tube.
- For deeper sheetflow, rely on gravity to deliver the sample.
- Use sampling diverters to direct and channel flow into sample containers. This may be particularly helpful in areas where shallow sheetflow is produced to generate a sufficient sample volume (e.g. 500-1000mL) per storm.
- To reduce sampling costs, samples from multiple source areas (of the same type) may also be combined into one sample before sending it to the lab for analyses.
- If using a peristaltic pump, the Teflon tubing may need to be changed or cleaned between each sample collected to prevent cross-contamination of samples (Burton and Pitt, 2002).

What Minimum Data is needed to Characterize Drainage Area/Site Conditions?

The relatively small size of source areas makes it very important to accurately determine the size of the contributing drainage area to each source area monitoring location. It is recommended that measurements of drainage area characteristics (e.g. size, land cover, location of inlets) be accurate to within a tenth of a foot to ensure that no other source areas are contributing to the sampling site. High-resolution imagery (e.g. 1m) may be used prior to site visits to delineate the drainage areas while site visits can be used to verify the catchment characteristics and confirm drainage area boundaries. Site visits during runoff events are particularly helpful in establishing drainage patterns that may not be evident from maps or imagery alone. Photographs accompanied by detailed notes on potential sites are very useful to determine sample collection device placement, and the number of sampling sites, and to aid in data analysis and interpretation.

Specific information recommended to characterize the most common source area types is given in Table 4. This information can be related to runoff characteristics. For all source area types, the following should be documented:

- Contributing drainage area,
- Land use and land cover,
- Impervious area and pervious area, noting any compacted pervious areas that may act like impervious areas,
- Rainfall,
- Topography (site slope),
- Any stormwater treatment practice on site,

- Runoff characteristics (based on selected monitoring parameters), and
- Soil types (given land use, soils may have reduced infiltration capacity post development use and may vary with USDA soil type classification, for example more compacted). Soil measurements from soil cores and soil penetrometers are suggested to provide up-to-date soil classifications.

Table 4. Site characteristics of four source areas	
Lawns	turf cover and quality, drainage area, soil conditions, soil phosphorus (P) index[1], soil compaction, fertilization history, runoff coefficient, interview homeowner about lawn management practices, lawn slope
Rooftops	roof type, age, downspout material, drainage area, roof pitch, presence of overhead tree canopy
Streets	street and gutter condition, traffic volume, presence/absence of street sweeping, deicing practices used, parking restrictions, on-street parking, presence/absence of street trees, density of vegetative cover, open or closed section drainage, closed drainage – number of outlets and locations, curb vs. grate opening (affect sampling equipment), presence/absence of catch basins
Parking lots	pavement conditions, slope, other land cover types such as landscaped medians, average parking density or parking lot use such as percent occupied

[1] Site characteristics are used to identify the potential phosphorus movement from a site. The ranking of Phosphorus (P) Index identifies sites where the risk of phosphorus movement may be relatively higher than that of other sites.

How Much Sampling Effort Is Needed to Get Reliable Data?

- Use data from previous studies or other published data on source areas to determine an approximate number of samples needed to get reliable data. Consult Appendix A for statistical methods to estimate sample size.
- In humid regions such as the northeast United States, it is likely that 10 good quality samples may be collected per year for each site (based on rainfall patterns). This can be used to determine the length of the monitoring study given the estimated sample size using methods presented in Appendix A. If monitoring occurs over different seasons, attempt to get a similar number of samples from each time period. An important note is that you may need to collect twice as many samples to get an acceptable number of 'keeper' samples given problems that may be encountered that require discarding an entire sample set.
- Review historical precipitation records to determine what constitutes a runoff event for the study area. Consult local forecasts to assess satellite and radar images to determine if the precipitation is likely to produce a runoff event. Storm depth will need to typically exceed 0.2 inches before sufficient runoff is produced but this will vary regionally and with land cover type. Determining which storm events to sample may use the following criteria as a starting point but in practice it is often based on experience from site personnel.

  - No greater than a trace of precipitation 24hrs preceding the storm event
  - Must be a runoff generating rainfall
  - Runoff must be sufficient to generate the required water volume for constituent analysis
- Document rain and flow conditions during the sampling period and for a short period of time before sampling. Pollutant concentrations vary with flow generated throughout a storm event. For example, describing the rain conditions prior to the first sample collected is important since rain conditions up to the time of sampling can have a significant effect on measured pollutant concentrations. Flow from these small source areas may be estimated using an empirical method.
- Collect multiple samples throughout the rain event. These discrete samples collected over a short period of time better characterize the source area than a single sample, with the exception of studies whose purpose is to determine first-flush characteristics.

What are Special Data Management and Quality Control Considerations?

It is recommended to follow the basic elements of a QA/QC plan for the study design provided in Section 2 and Appendix G of this manual. In addition, the following are specific considerations for monitoring source areas.

- Open containers (common to source area sampling) are prone to contamination and therefore field personnel should load pre-treated bottles into collection device shortly before the precipitation event
- Maintenance of sampling equipment may be more difficult when samplers are located in hard-to-reach locations (e.g., in a catch basin), so you may need to allow more time for this.
- Pre-test samplers to determine if the sampling equipment is functioning properly (e.g. bottles not overfilling, debris jams, missed flows etc.). Modifications of the sampling set-up or equipment may be needed based on these preliminary results.
- Coordination of sample collection can be difficult when monitoring multiple source areas for a single event. Mapping out efficient sample retrieval routes can help to ensure the minimum sample handling times are met.

How Much Should be Budgeted for the Source Area Monitoring Study?

An example project budget with a sample narrative for source area monitoring of lawns is provided in Box 1 and Table 5. Sampling costs vary greatly based on the number of parameters analyzed and whether an in-house laboratory is available. The use of volunteers may also reduce costs, but would require additional training costs. Project managers should include some buffer in the budget for equipment costs to account for replacement given the potential for vandalism or large storm events that may damage equipment.

Box 1. Description of lawn fertilization study to support example budget

The purpose of this study was to determine the impact of residential lawn fertilizer applications on stormwater runoff quality in residential neighborhoods. Ten residential lawns, stratified by fertilizer rate, were monitored for 20 storm events over a 2-year period. In addition, two control sites were selected: a forested site and an agricultural site within the watershed. Parameters for analyses included: total nitrogen, inorganic nitrogen and organic nitrogen, sediment solids, and total and dissolved phosphorus. The sampling methods are based upon Waschbusch et al. (1999).

A residential lawn care survey was completed to identify lawns with high, low and no fertilizer application rates (See Profile Sheet 5 for information on survey methods). This information was entered into a GIS database and overlain with parcel data to randomly select lawns to be included in the study. At the time of the survey, participants were asked if they would like to participate in the runoff monitoring part of the study.

The lawn samplers were installed in early Spring at all 14 sampling sites (12 lawns, 2 control) and located in a canopy-free area of the lawn. Prior to their installation, a topographic assessment of each lawn is made to determine the best placement of the sampler (e.g. at least 5% slope, sufficient slope length (20ft) and drainage area captured was free of objects that may obstruct runoff flow, or collect runoff from other source areas) (Garn 2002). Small fluorescent flags were installed to demarcate the location of the sampling equipment to prevent trampling on the equipment. Soil samples were taken from each lawn to determine the soil quality and condition (e.g. pH, texture, density, nitrogen, phosphorus and carbon content). The lawn management survey provided information on land cover characteristics and nutrient inputs from fertilizer.

Table 5. Example budget for residential lawn fertilization source area monitoring study			
SOURCE AREA SAMPLING	**Staff Resources**	**Unit Cost[1]**	**Total Cost**
Monitoring 12 sites (10 lawns, 2 control), 20 storms			
PLANNING (25%)			
Background Research (incl. data acquisition)	40 hrs		$2,000
Desktop analysis	32 hrs		$1,600
Field reconnaissance for final site selection (incl. homeowner interview and permissions)[2]	80 -100 hrs		$4,000 - $5,000
Project scope and sample design	40-80 hrs		$2,000-$4,000
Develop monitoring plan	40 hrs		$2,000
PLANNING SUBTOTAL			**$11,600 -$14,600**
IMPLEMENTATION (75%)			
Equipment and supply costs [3](e.g., latex disposable gloves, sample bottles, sample collection device, coolers for sample storage)			$6,250
Training (Staff and/or volunteers)	3 day, 2 staff		$1,600
Sample collection, storage and transfer[4]	240 hrs		$12,000
Sample analyses 5 (TSS, BOD, TP, TN, TKN, NO2, NO3)		$120	$14,400
Data analysis and interpretation	80 hrs		$4,000
Final Report	80 hrs		$4,000
IMPLEMENTATION SUBTOTAL			**$36,250**
TOTAL			**$53,850 - $56,850**

[1] Assume $50/hr
[2] Allows about 1-hour per site to include travel
[3] Will vary based on method (e.g. grab bottles to complex sampler design), assume a 25% replacement cost
[4] 20 samples, collected per site. Allows 1-hour per site to included travel, site maintenance, rainfall measurements
[5] 10 of the 20 samplers are "keeper" samples, see Appendix C for cost estimates

What Monitoring Problems Can Be Anticipated?

- An insufficient number of representative samples may be collected due to a variety of reasons:
  - Insufficient rain events that meet sampling criteria. This can result in a need to extend the length of the monitoring program.
  - It is often difficult to get an uncontaminated full sample from every storm or site and may require many more sampling attempts than originally estimated to collect sufficient representative samples.
  - Sampler bottles may over fill and act as a sediment trap. It may be necessary to discard an entire set of source area samples for that particular event.
- Damage to in-situ samplers can occur given their visibility or location in high traffic areas (both pedestrian and vehicular). Check and maintain equipment in

between and just after sampling events. Have additional supplies available so that repairs or equipment replacement can be made quickly in the field.

- Availability of personnel to monitor, reload and retrieve bottles can be a limiting factor to collect sufficient data. The study design should address if resources (personnel, funds) are available to sample at all times (e.g. 24 hrs/day, 7 days/week) versus Monday through Friday. Finding cooperative property owners and the use of an effective volunteer program may help alleviate this issue.
- Changes in the drainage area can occur during the monitoring study such that runoff from a separate source area is directed into the study site. In this situation, detailed land use and land cover data that was collected during source area site selection may be used to estimate this additional contributing runoff and its characteristics to this expanded drainage area.

What are Some Good Monitoring Resources to Consult?

New Zealand rooftop runoff study http://www.kml.co.nz/proj10RoofRunoff.htm

Marquette, MI study to quantify sources of contamination in an urban basin http://onlinepubs.er.usgs.gov/djvu/WRI/wrir_97_4242.djvu

Madison, WI study to identify sources of phosphorus in stormwater in two urban residential basins http://onlinepubs.er.usgs.gov/djvu/WRI/wrir_99_4021.djvu

City of Austin, Stillhouse Spring Cleaning, available on-line at http://www.ci.austin.tx.us/growgreen/stillhouse.htm (accessed 6/25/07)

How To Do Stormwater Sampling - A guide for industrial facilities by Steven Golding and Norm Glenn. http://www.ecy.wa.gov/biblio/0210071.html

Stormwater Effects Handbook. Chapter 5. http://unix.eng.ua.edu/~rpitt/Publications/Publications.shtml

Weather data

NOAA's National Weather Service http://www.nws.noaa.gov

The Weather Channel http://www.weather.com/

Design

3

Performance Monitoring of Individual Stormwater Treatment Practices

What is the Monitoring Study Design?

MS4 stormwater managers have a critical need to monitor stormwater treatment practices (STPs), as there is much uncertainty in their perceived as well as their actual documented performance. Currently, three primary sources of information on STP performance are available that report on the pollutant removal efficiencies for the majority of STP types. These include a summary of 166 STP performance studies published in National Pollutant Removal Performance Database (CWP 2007), results from over 300 STP studies presented in the International Stormwater Best Management Practices (BMP) Database (http://www.bmpdatabase.org) and a web-based tool supported by the U.S. EPA that provides access to studies covering a variety of traditional and low impact STP (http://www.epa.gov/npdes/urbanbmptool). These data sources illustrate the wide degree of variability in pollutant removal efficiencies based on the type, location, and design of the STP, in addition to how the data is collected and analyzed, and the natural variability of stormwater characteristics.

These databases primarily include non-proprietary STPs. In terms of proprietary treatment devices, performance monitoring is expensive and generally not available. However, some manufacturers and government agencies have formed groups to establish uniform protocols to test these devices and address the performance-related concerns of stormwater managers (Refer to Appendix F for further information).

Due to this variable and limited performance data, MS4s need additional performance data to identify effective STPs for reducing stormwater runoff volume and directly target their pollutants of concern. This study design presents a standard approach to monitor the performance of individual structural STPs to provide results that are more comparable locally, regionally and even nationally. The approach is most useful to collect data for STPs that are new or innovative, or to test improved versions of older STPs. This study design does not apply to proprietary devices due to the enormous range and variability of these practices. Results generated from implementing this study design may be submitted to the International Stormwater BMP Database and the Center for Watershed Protection to further the existing knowledge base.

For the purposes of this study design, STPs are categorized into simple and complex monitoring situations. Simple STP monitoring situations occur when the structure is relatively small and has a defined inlet and a defined outlet, where both runoff volume and quality can be effectively characterized (e.g. a wet pond). Complex STP monitoring situations occur when flow into or out of the structure is distributed and cannot be effectively characterized without concentrating or redirecting flow, or making additional measurements or estimates of soil water quality, groundwater movement, or runoff volume (e.g. bioretention facilities, swales). Examples of simple and complex monitoring situations are illustrated in Figure 1.

Figure 1. a) Simple STP monitoring situation – defined single inlet and single outlet of a wet pond; b) complex STP monitoring situation – multiple inlet locations, evapotranspiration and infiltration losses at a bioretention facility

Why is Performance Monitoring of Stormwater Treatment Practices needed?

The use of a standard monitoring design to monitor the effectiveness of individual STPs can accomplish the following management objectives:

- Document how well a community is meeting pollutant reduction goals.
- Evaluate if pollutant removal in the community differs from national estimates.
- Evaluate the runoff reduction that may be achieved by specific STP types.
- Provide new or updated pollutant removal rates for standard stormwater practices for specific water quality parameters of local interest (e.g., bacteria, pesticides, hydrocarbons, trash, and debris).
- Develop new removal rates for untested or innovative practices.
- Test whether specific design features can boost reported median pollutant removal rates or enhance runoff reduction volumes.
- Evaluate removal rates under seasonal or special conditions (spring snowmelt, high or low sediment loadings).
- Provide scientific support for enhancing stormwater design criteria or removal rates within the local manual.

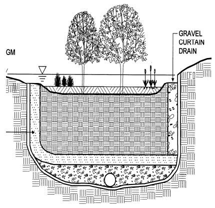
- Define the degree of stream warming produced by the STPs.
- Provide opportunities to share data and coordinate monitoring efforts at regional scales.

What is the Basic Approach?

The basic monitoring design is a longitudinal study design where paired sampling of inflow and outflow are taken for either simple or complex monitoring situations. Accurate measurements of flows into and out of the STP along with accurate site rainfall measurements are critical to estimating the performance of the STP. These runoff and rainfall measurements are the largest sources of error and are the main differences between simple and complex STP monitoring situations. The design of simple STP monitoring studies is more straightforward compared to complex STP monitoring programs because of the ability to sample the well-defined inlet and outlet structures for flows. There are also a substantial number of prior monitoring studies for simple STPs, which can be learned from and incorporated into the design of the monitoring program. Complex STP monitoring situations typically require a more specialized sampling strategy, as presented below.

Complex STP monitoring situations are more difficult to sample because stormwater enters many of these practices as sheetflow or through multiple inlets and often exits through infiltration processes. Table 1 provides a few example techniques to consider monitoring the flow into a complex STP while Figure 2 illustrates an example monitoring set-up for a bioretention practice. The study design for monitoring complex STPs should be modified to account for losses, such as groundwater and evapotranspiration. Although these losses are not explicitly monitored and are difficult to measure, they are essential to enhance the pollutant removal and flow reduction capabilities of complex STPs. Estimates of these losses from complex STP monitoring situations may also be accomplished by using a paired site monitoring approach, where one site acts as a control and the other as a treatment. Paired site monitoring can include one site with a complex STP and one without, or it can include two sites with variations in complex STPs as a way to monitor losses that are difficult to account for by looking at the differences in outflow, for example between the control and treatment sites. For example, one site could have a lined bioretention facility with an impermeable membrane to eliminate exfiltration, while another site could have an unlined bioretention facility to allow both exfiltration and evaporation. With this approach, the flow lost to exfiltration can be estimated to better evaluate performance of the STP. The difficulty associated with this type of sampling approach has resulted in fewer studies conducted for complex STP monitoring situations. Therefore, the study design will generally be unique and need to account for site-specific conditions.

For both simple and complex monitoring situations, a phased approach to monitoring may find that the selection of sampling sites and the collection of field data may need to be modified after initial data has been collected. These initial data are also needed to confirm the experimental design assumptions in terms of sample size, with some modifications likely necessary.

Table 1. Four Options for Sampling Distributed Sheet Flows into a Complex STP Monitoring Situation	
Weir  Source: Smith et al., (No Date)	Estimate flow by installing a weir or sump at the inflow
Underdrain  	Install an underdrain to collect runoff that would have otherwise infiltrated
Runoff Estimation **$R = P*Pj*Rv$** **Where:** **R = annual runoff (inches)** **P = annual rainfall** **Pj = fraction of annual rainfall events that produce runoff (usually 0.9)** **Rv = runoff coefficient (0.05 + 0.9I)**	Measure outflow and estimate inflow using an empirical runoff estimation method such as the one shown

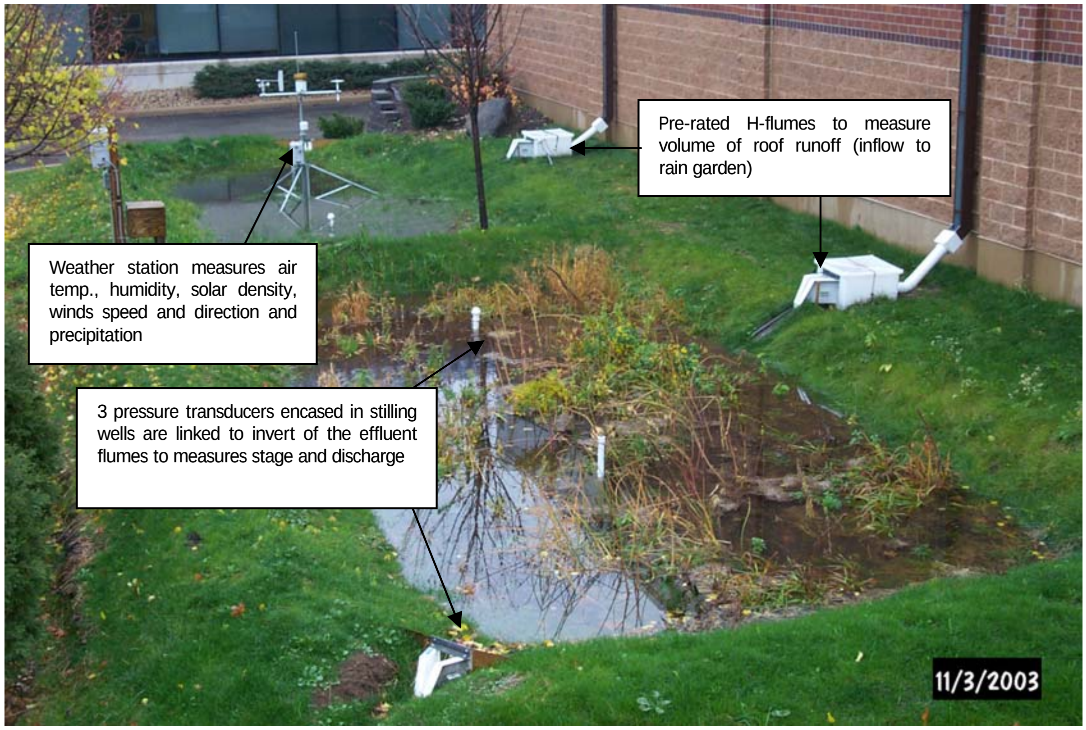


Figure 2. Example monitoring set-up to estimate a water mass balance for a paired rain garden study. (Courtesy of W. Selbig, USGS-WI)

STP performance can be measured in terms of pollutant removal or runoff reduction. The two predominant methods of determining pollutant removal effectiveness are mass efficiency, also referred to as "summation of loads" and event-mean-concentration (EMC) efficiency**,** also referred to as "efficiency ratio". Table 2 provides a description of these two calculations, as well as the data required for each. Comparisons of a wider range of methods are presented in CWP (2008), ASCE and EPA (2002), Burton and Pitt (2002) such as the regression of loads, mean concentration and efficiency of individual storm loads. The mass efficiency approach is recommended because it is influenced by the volume of water entering the STP and water losses within the STP (e.g., evapotranspiration and infiltration). As a result, a water mass balance can be calculated to assure that inflows and discharges have adequately accounted for any gains or losses.

It is critical that performance is not only described by a single value. Most studies have shown that STPs have varying pollutant removal efficiencies that are affected by flow rate (or volume) and influent concentration. If a single value is given, it is only applicable for the site where it was determined and for a very similar series of storms monitored. As a consequence, STP performance data should be reviewed with a careful eye, keeping in mind the conditions for which the value may be appropriately interpreted. As part of the study design, it is recommended to report and monitor a range of storm events and provide detailed characteristics about the STP design factors and contributing catchment areas. These measures of efficiency are valid only if a sufficient number of representative samples are used in its calculation. Refer to *How*

Much Sampling Effort is needed to get Reliable Data and Appendix A for information on the number of samples and monitoring sites needed.

The amount of runoff reduction associated with an STP is rarely quantified yet integral to the effectiveness of many SPTs. STPs that reduce volume are also reducing pollutant loads, although a concentration-in vs. concentration-out study would not account for this reduction. For this reason, the pollutant removal efficiency of these types of STPs may be under-reported. This is perhaps the most crucial factor for complex STP monitoring situations. As previously mentioned in the study design, if total effluent quantity cannot be accurately measured, paired site monitoring at a similar site without a control should be considered. Measurements of runoff reduction enhance the degree and reliability of pollutant mass removal for STPs because pollutant loads are the product of both stormwater flow volume and the treated pollutant concentration leaving the practice.

Every monitored site is idiosyncratic with respect to its contributing drainage areas, sampling intensity or quality, STP history, STP design features, incoming stormwater characteristics, and the number and types of storm events sampled. No single STP monitoring effort can define the expected removal rate for a practice or draw sound conclusions that is applicable everywhere. However, it can help to draw strong inferences by contributing to the growing body of performance studies and comparing results.

Table 2. Methods of Determining STP Effectiveness

Method	Calculation SOL = sum of pollutant loads[1]	Data Needs
Mass Efficiency	[(SOLin-SOLout)/SOLin]*100	flow, precipitation, pollutant concentration
EMC[2] Efficiency	[(EMCin-EMCout)/EMCin]*100	pollutant concentration

[1] SOL is the sum of the product of the EMC and total storm volume for each event
[2] EMC is a statistical parameter used to represent the flow-proportional average concentration of a given parameter during a storm event. It is defined as the total constituent mass divided by the total runoff volume. When combined with flow measurement data, the EMC can be used to estimate the pollutant loading from a given storm.

What Factors should be Considered when Selecting Test Sites?

The goal in the selection of test sites is to select STPs that are generally representative of the population of STPs in the community. Monitoring representative STPs can allow results to be applied on a broader scale. Desktop analyses of potential monitoring sites and field reconnaissance are both needed to screen sites using a list of pre-defined criteria. At a minimum, the criteria should include STPs with design features that optimize their function rather than selecting STPs that may have an 'expected' poor performance given their poor design features, or lack thereof. Other criteria include the ability to collect accurate data, and safety of the monitoring crew. When monitoring stormwater treatment practices, sampling locations need to be identified to accurately characterize influent and effluent flows. With complex treatment systems, paired site monitoring may be needed. However, collecting samples as near to the STP as practical reduces the likelihood that other factors will influence the representativeness of the data. When in-stream measurements are being taken for compliance monitoring to indicate the in-stream benefits of the stormwater controls, additional and complex considerations are needed. Burton and Pitt (2002) review the tools and processes in detail for in-stream monitoring. Some additional practical factors include:

- Access and use of electricity,
- Single inlet and outlet for simple STP monitoring situations,
- Good and safe access for crews,
- Sufficient space to locate and install equipment,
- Location is safe from vandalism and physical damage from storm events and is not susceptible to erosion etc.,
- Landowner permission,
- Available data from previous monitoring efforts, and
- Known representative influent EMCs.

The selection of monitoring site(s) is a potentially lengthy process in the monitoring program. It is expected that only a small proportion (i.e. 1 out of 20) of sites examined for a representative stormwater control will meet the pre-defined selection criteria.

Identifying good candidate sites for monitoring provides some assurance that quality data may be generated to achieve the monitoring objectives (e.g. evaluate the pollutant removal efficiencies of the STP). One of the most important factors in the selection process is whether the proposed STP represents the best possible set of design features that can be achieved in the community. It is impractical to document poor removal by a poorly designed STP; however, compliance monitoring in a community must consider typical applications of the practice. While many STPs may not be constructed with the optimum design features, research and field reconnaissance can validate that the sites chosen are most representative of the highest standard constructed in the community. Table 3 below provides design features that enhance and diminish the performance of swales as an example of factors that can be used to accept or reject test sites. Additional examples of design features that enhance and diminish the performance of other STPs can be found in *Hirschman et al.* (2008).

Table 3. Example Factors that Enhance or Diminish Performance of Dry Swales

Enhancing Factors	Diminishing Factors
Exceeds target WQv	Does not provide full WQv
Turf cover with trees, shrubs or herbaceous	Turf cover only
Longitudinal swale slope between 0.5 to 2.0%	Longitudinal swale slope <0.5% or >2%
Measured soil infiltration rates exceed 1.0	Measured soil infiltration rate less than 1.0
Off-line or multiple treatment cells	No pretreatment to the swale or channel
Lacks underdrain or uses underground stone	Swale sideslopes more than 5:1 h:v
Media depth more than 24 inches	Intersects groundwater (except wet swale)

What Unique Sampling Techniques and Equipment are needed?

The minimum sampling equipment needed to generate precipitation, flow, and water quality data includes: rain gage, flow meter, and automatic sampler. A digital camera is also recommended for photo documentation of the facility. Rain and snow gages, where appropriate, should be installed as close as possible to the monitoring stations because in many regions precipitation is highly variable within a small area. Manual rain gages should also be used at the monitoring site to check accuracy, consistency, and proper functioning among different gages (ASCE and EPA, 2002).

Automated samples are recommended, as opposed to manual grab samples, for both simple and complex STP monitoring situations. Automated sampling provides accurate collection of flow-weighted composite samples, provides the ability to include early flows in the sample, and eliminates the need for an operator to be on-site during stormwater sample collection. With the accurate collection of flow data, the mass efficiency approach can be used to determine the effectiveness of the STP. It is important to have routine inspection and maintenance on automatic samplers to help ensure that the equipment will function properly when a storm event occurs (ASCE and EPA, 2002). Although automated sampling is the recommended approach, it cannot be used for some water quality parameters such as bacteria, oil and grease or volatile organics. Manual sampling is required to avoid chemical transformations of the sample (sterile equipment, minimize storage time) for these parameters.

Additional equipment that may be useful if resources are available includes a water quality probe and a bedload sampler. The bedload sampler collects bed load material (greater than about 250 μm particles), which may be important when characterizing stormwater sediment discharges. This can provide a missing component of the mass balance in stormwater studies (Burton and Pitt, 2002).

What Minimum Data is needed to Characterize Drainage Area / Site Conditions?

There is often a significant relationship between influent and effluent concentrations for various parameters based on the type of STP and how sensitive the discharge concentration is to the influent value (Lampe et al., 2005). This relationship may also be a factor for selecting monitoring parameters and how they will impact the calculated effectiveness of the STP. Typical water quality monitoring parameters include:

- Nutrients: Nitrogen and Phosphorus
- Particulate Matter: Total Suspended Solids
- Toxicants: Copper, Lead, and Zinc
- Pathogens: Fecal Coliform
- Other: Hydrocarbons

The following additional information is needed for characterizing the drainage area and site conditions:

- Detailed description of the contributing drainage area to include: catchment size and boundaries, land use and land cover types, development characteristics (e.g., age, traditional versus environmental site design), type of conveyance system (e.g., open or closed channel) leading to inflow point.
- Basic characterization of inflow EMCs to the STP to confirm the sample size estimate that is based on the variability of EMCs (see Appendix A). A few months may be needed to collect enough samples to determine whether the runoff to the STP is stronger or weaker than expected based on published values. The sample size may be needed to be adjusted accordingly.
- Rainfall data (on-site is preferable to proximal rain gage) to derive runoff coefficients and relate EMCs to storm size. Rainfall data also helps to determine when to start sampling. Consideration for more than one rain gage per study site for larger drainage areas or in smaller ultra-urban catchments when storm variability can be significant at these scales.
- Historical rainfall data to determine the frequency of different size storm events, on which to base sample collection
- STP engineering data to define the geometry and sizing volumes employed, such as inlet and outlet dimensions and details, forebay length and volume, detention surface area, average daily baseflow volume, residence time. Urbonas (1995) provides a detailed list specific to various types of structural STPs.
- Operational history of the STP (age, maintenance, nuisance problems, such as geese, changes in vegetative cover, etc.).

How Much Sampling Effort is needed to get Reliable Data?

The sampling effort includes both the number of samples and the type of storms to be sampled for this study design. It is reasonable to expect to collect up to 10 paired storm event samples per year.

The desired number of samples should be statistically determined based on the parameter of interest. Appendix A contains more detail on the statistical analysis for determining sample size. In general, more paired samples are needed to determine a significant difference between the influent and effluent the greater variability there is in the inflow and outflow concentrations. An example in Table 4 shows the sample size needed to achieve a confidence level of 95% based on the percent difference in inflow and outflow means, typical sample concentrations (coefficient of variation of about 1), and a power of 80%. The sampling effort becomes more intensive as the difference in influent and effluent concentration becomes smaller (e.g., sampling efforts increase from 20 to 300 samples as the difference between influent and effluent decreases from 80% to 20%) – harder to detect smaller differences. This could mean the difference between a 1 or 5-year data collection period. Criteria must be set for determining when the pollutant removal efficiency is significant to determine if/when additional monitoring is required. Once the sample size has been determined, an iterative process to re-scope the project to remain within budget and collect a reasonable sample size may be needed.

Table 4. Number of paired samples needed based on the difference in inflow and outflow means (confidence level = 95%, power = 80%, coefficient of variation = 1)

Difference in Sample Set Means	80%	60%	40%	20%
# Samples Needed	20	50	75	300

It is important to collect samples from different size storm events since pollutant concentrations vary with storm flows. For example, Figure 3 is a line plot that illustrates the performance of a STP under different flow conditions with varying influent suspended sediment concentrations (500 mg/L, 250 mg/L and 50 mg/L). For example, this plot serves more as a visual aid to understand how STP performance may vary under different flow conditions. There is a much greater reduction in suspended sediment concentrations for higher concentrations (500 mg/L and 250 mg/L) compared to lower concentrations under similar flow conditions. Depending on the monitoring objective, it may also be important to characterize the water quality of dry weather flows for some STPs with significant wet storage and/or baseflows (ASCE and EPA, 2002).

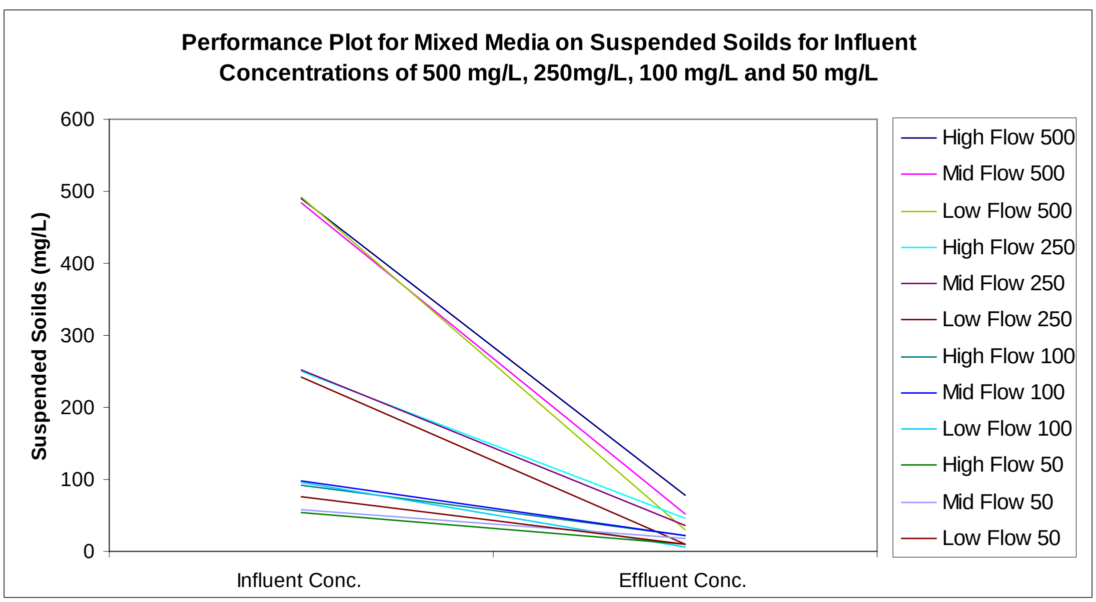


Figure 3. Performance plot for mixed media for suspended solids at influent concentrations of 500 mg/L, 250 mg/L, 100 mg/L and 50 mg/L

Samples should be collected during the complete storm event to obtain flow-weighted pollutant concentrations. It is suggested that composite samples be based on sub-samples through the entire storm event rather than a volume-based approach. These composite subsamples provide better results than samples collected during the first 30 minutes of the event, often referred to as the first flush (Maestre et al., 2004). First flush is dependent upon numerous site and rainfall characteristics, mostly the amount of pavement in the drainage area, the size of the drainage area, the complexity of the drainage system, and most importantly, the distribution of very high rain intensity periods. The use of a water quality sonde provides data every several minutes for important water quality indicators (turbidity, conductivity, temperature, DO, pH, etc.). These data can be used in compositing the water quality samples to represent obvious periods of the rain likely having very different characteristics.

For devices that have significant storage (such as wet detention ponds), simultaneous inflow and outflow samples may not be actually paired. As an example, during small events, the effluent water from these STPs may mostly be water that has been stored and displaced by a relatively small volume of incoming water. In this case, seasonal statistics using sum of loads and not concentrations are needed in contrast to individual event concentration data.

What Special Data Management and Quality Control Issues can be expected?

It is recommended to follow the basic elements of a QA/QC plan for the study design provided in Section 2 and Appendix G of this manual. In addition, the following are specific considerations for monitoring individual STPs.

As a result of sample collection and laboratory analysis, the STP monitoring program generates a considerable amount of information in a wide variety of forms. Both hard copy and electronic information needs to be stored, retrieved, and transferred. A central file is recommended to accommodate hard copy information and a database to accommodate digital information (ASCE and EPA, 2002).

Although the mass efficiency (SOL) method is a recommended approach to measure the pollutant removal effectiveness of STPs, it is challenging to obtain good paired data on flow and concentration. As a result, there may be a tendency to rely on historical EMC-based methods such as the "efficiency ratio" (see ASCE and EPA 2002, CWP 2008) However, Figure 3 illustrates how a single measurement of performance may be misleading by averaging individual storm events. High influent concentrations, as illustrated in Figure 3, result in large percentage reductions, while low influent concentrations result in much lower effluent concentrations. Similarly, if an influent is comprised of a large fraction of coarse, or large-size particles, the percent removal would also be artificially elevated compared to a sample with a more representative distribution of sediment-size particles. Using a mass efficiency method representing a long series of events representative of a range of storm events is preferable to a simple percent removal using EMCs based on individual events. Alternative approaches are presented by ASCE and EPA (2002) that use a variety of statistical methods to more fully describe, the inflow and outflow water quality of the STP to determine if significant differences exist.

Influent concentrations need to be reviewed to identify potential irreducible concentrations that typically exist for most STPs and may affect how STP efficiency estimates are interpreted. Irreducible concentrations represent the lowest concentration that may be achieved by an STP and the effluent quality discharged from an STP that represents the best that can be achieved. They exist due to internal processes within the STP that inevitably return some pollutants back into the water column or simply reflect the limitations of a particular removal pathway utilized in the STP. When irreducible concentrations are a factor, it may be more useful to report the efficiency of the STP relative to the achievable level of treatment (Schueler, 2000; ASCE and EPA, 2002). In all cases, effluent quality achieved by the STP should be reported.

Paired box plots of influent and effluent quality are also useful as a visual analysis. They reveal that effluent quality is much less variable than the percent of pollutant removed. These box plots typically present the median, the upper and lower 95 percent confidence levels of the median, and the 25th and 75th percentiles (Strecker et al., 2004). Figure 4 presents an example box plot of the copper influent and effluent of a bioretention facility.

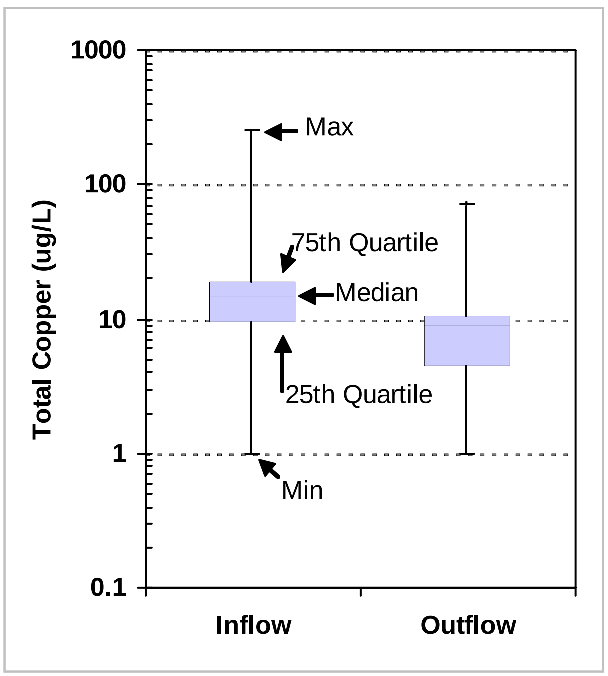


Figure 4. Box plot for bioretention copper influent and effluent concentrations

Other factors to consider include:

- Consistent sampler timing among all automated equipment,
- Check that runoff volumes in and out are present,
- Collect an equivalent number of samples for inflow and outflow, and

- Pollutant removal rates should not be calculated unless at least 5 to 10 paired storm events are available for analyses that meet the operational criteria and the variation of the samples can be determined.

How much should be budgeted for the Monitoring Study?

An example of a budget for monitoring structural STPs is provided in Table 5 for both simple and complex STP monitoring situations. This table provides general guidance only, and will vary with equipment used, monitoring parameters, site constraints, in-house vs. contract lab analysis, and other factors. The basic approach for both types of monitoring situations is to install automatic samplers at the inflow and outflow. The complex situation also involves installation of a flow concentrator (e.g ., flume or weir) to direct flow to an automated sampler at the inlet and underdrain to measure the outflow. The budget in Table 5 assumes that a total of 30 storm events will be monitored over a two-year period, and that 20 will meet QA/QC standards. It also includes collection of samples to characterize baseflow conditions. Other assumptions include: samples are sent to a contract lab and analyzed for standard pollutants/constituents, including nitrogen, phosphorus, total suspended solids, lead, zinc, fecal coliform, and hydrocarbons.

Table 5. Individual Structural STP Monitoring Budget for Simple and Complex Situations						
	Simple STP Monitoring Situation			**Complex STP Monitoring Situation**		
	Staff Resources	**Unit Cost**	**Total Cost**	**Staff Resource s**	**Unit Cost**	**Total Cost**
Planning	**5%**			**6%**		
Background Research (identify potential STPs, determine data needs and monitoring parameters)	40 hours	$50/hour	$2,000	40 hours	$50/hour	$2,000
Desktop Analysis (major tasks include: preliminary site selection, preliminary site characterization, generate field maps)	32 hours	$50/hour	$1,600	32 hours	$50/hour	$1,600
Field Reconnaissance and Site Selection	32 hours	$50/hour	$1,600	32 hours	$50/hour	$1,600
Project Scope and Sample Design	16 hours	$50/hour	$800	32 hours	$50/hour	$1,600
Develop Monitoring Plan	8 hours	$50/hour	$400	16 hours	$50/hour	$800
PLANNING SUBTOTAL			**$6,400**			**$7,600**
Implementation	**95%**			**95%**		
Equipment [1]			$15,000			$17,000
Equipment Installation and Maintenance[2]	256 hours	$50/hour	$12,800	512 hours	$50/hour	$25,600
Training	32 hours	$50/hour	$1,600	32 hours	$50/hour	$1,600
Sample Collection[3]	512 hours	$50/hour	$25,600	512 hours	$50/hour	$25,600
Sample Storage and Transport			$10,000			$10,000
Chemical Analysis[4]		$200 per sample	$8,800			$8,800
Data QA/QC	40 hours	$50/hour	$2,000	40 hours	$50/hour	$2,000
Data Analysis and Interpretation	80 hours	$50/hour	$4,000	80 hours	$50/hour	$4,000
Final Report	80 hours	$50/hour	$4,000	80 hours	$50/hour	$4,000
IMPLEMENTATION SUBTOTAL			**$83,800**			**$98,600**
TOTAL			**$90,200**			**$106,200**

[1]Simple = 2 automatic samplers, triggering sensors, pump, lumber, concrete, battery, waders, clipboards, fieldbooks, first aid kits Complex = 2 automatic samplers, triggering sensors, pump, lumber, concrete, battery, pipe for underdrain, flow concentrator at inlet

[2]Maintenance for simple assumes 1 person, 2 hours per week, for 2 years. Maintenance for complex assumes 1 person, 4 hours per week, for 2 years. Installation for simple assumes 3 people for 2 days. Installation for complex assumes 3 people for 4 days.

[3]Sample collection assumes 2 people for 8 hours for each storm event. A total of 30 storm events will be sampled and 2 baseflow events. Out of the 30 sampled events, only 20 are expected to meet QA/QC standards.

[4]Chemical analysis assumes contract lab analysis for standard pollutants/constituents. One composited inflow and one composited outflow sample will be analyzed for a total of 20 storm events and 2 baseflow events.

What Monitoring Problems can be anticipated?

Some common problems encountered during STP monitoring studies:

- An overly sensitive trigger for sampling equipment could lead to mistakenly collected samples
- Problems related to seasonal responses (e.g., drought conditions, heavy storms that destroy equipment)
- Miscalculating storm size and under or over-filling bottles
- Site access is important for field employees, but can also lead to vandalism
- Total inputs and outputs are not accounted for, such as gross pollutants missed by automated sampler. Closing the mass balance is important for determining mass efficiency.
- Equipment failure
- Failure to synchronize the two samplers over the same hydrograph
- Storms produce inflow but no outflow.

Incorporating a rigorous maintenance schedule, training, and practice runs of sample retrieval may help to alleviate these problems. Adjusting expectations and building time and money into the budget to address potential problems is also recommended.

What are some Good Monitoring Resources to Consult?

Burton, A., and R. Pitt. 2002. *Stormwater Effects Handbook: a Toolbox for Watershed Managers, Scientists, and Engineers.* Lewis Publishers. New York, NY.

American Society of Civil Engineers (ASCE) and United States Environmental Protection Agency (EPA). 2002. *Urban Stormwater BMP Performance Monitoring; a guidance manual for meeting the national stormwater BMP database requirements.* EPA-821-B-02-001 http://www.bmpdatabase.org/docs/Urban%20Stormwater%20BMP%20Performance%20Monitoring.pdf

Attachment SD3-A. Bioretention Case Study

The following steps describe the development of a hypothetical bioretention monitoring plan based on procedures developed for Charlotte, North Carolina (*Smith et al.*, no date), as well as procedures utilized by several professionals with expertise in bioretention monitoring (*Davis, pers. comm.*, 2006; and *Glass, pers. comm.*, 2006). The values used in this study are for illustration purposes only.

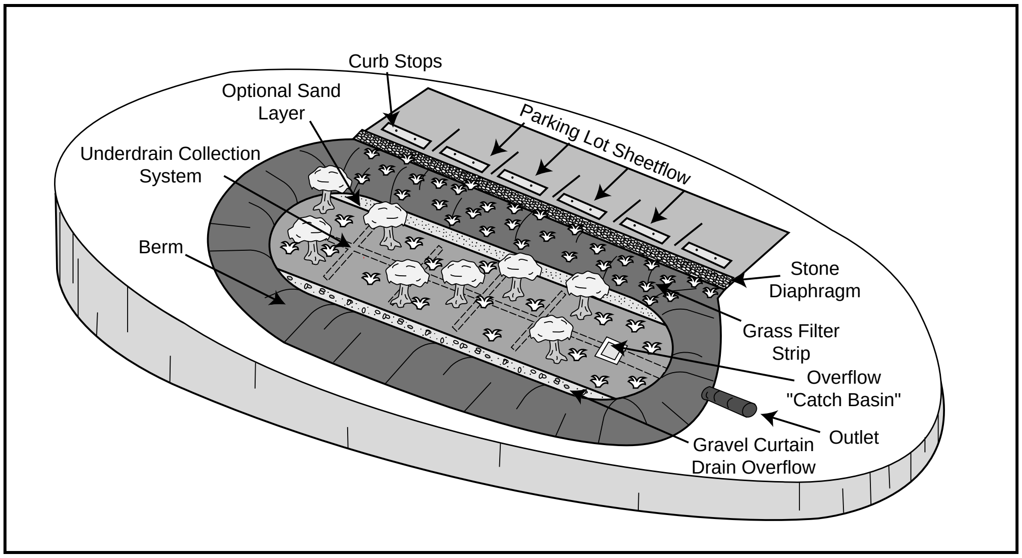


Figure SD3-A1. Isometric view of a bioretention facility

Management Issues to be addressed

- What is the effectiveness of bioretention practices in parking lots for typical pollutants found in stormwater in the region?
- Does the bioretention practice reduce concentrations of these pollutants to what are deemed acceptable levels?
- What is the runoff reduction capability of the bioretention practice?

Total # of sites investigated = 15

Study Design

Paired inlet and outlet sampling achieved by concentrating in the inflow and outflow.

Data to characterize drainage area / site conditions

- Obtained from the County: GIS data, including orthophotos, streams, land use/land cover, topography, and percentage of imperviousness.
- Obtained from the contractor hired to construct the facility: as-built drawings
- Obtained from USGS: streamflow and precipitation data

Water quality monitoring endpoints

- Nitrate-nitrogen
- Copper
- Lead
- Zinc
- Fecal coliform

Sampling effort

- 2-year
- 1 site met criteria out of 15 surveyed
- 20 storm event sample pairs, with a frequency of about 1 storm per month (Assuming a coefficient of variance of 1, power of 80%, and difference in the mean inflow and outflow concentration of 80%, this sample size will provide a 95% confidence level).
- Precipitation events monitored ranged from 0.1 to 2 inches of rainfall – in this region, precipitation events larger than 2 inches occur only a few times annually, and events less than 0.1 inches generally produce little or no runoff.

Sampling technique and equipment

Sheetflow normally enters the bioretention facility from a parking lot. In order to measure inflow, the parking lot was modified so that the sheetflow converged to a single inlet consisting of a 120 degree v-notch weir. Outflow was also directed to a single outlet consisting of a 30 degree v-notch weir. An ISCO model 6712 sampler was installed at both the inlet and outlet. Samplers were equipped with 24 1,000 mL sample bottles. The samplers were triggered during storm events and collected samples at 5-minute increments. One 200 mL sample was collected every 5 minutes; therefore, 5 composite samples were obtained in every 1,000 mL bottle. An ISCO model 720 bubbler flow meter was also placed at the inlet and outlet to monitor flow at 5-minute increments. To measure precipitation, an ISCO model 673 tipping bucket rain gage was installed slightly upstream of the bioretention facility in an area unhindered by overhanging trees and powerlines. The equipment and bioretention facility were inspected twice a week.

Measurements of fecal coliform were collected by manual grab sampling at the inlet and outlet with 250 mL sample bottles every 15 minutes. During the end of the storm event, the grab samples were composited. The composite samples from the ISCOs were also collected, placed on ice, and sent to a contract lab to be analyzed for the water quality monitoring endpoints. The process of sample collection, storage, and transport to the contract lab followed chain of custody guidelines set forth in the QA/QC plan.

Budget = $125,000

Data Analysis and Evaluation

- Mean Concentration and Summation of Loads methods were chosen to calculate the pollutant removal efficiency of the bioretention facility.
- Nonparametric statistics were used to compare the influent and effluent concentrations. The Mean Concentration analysis of influent and effluent for this study is summarized in Table 6. Average inflow and outflow concentrations are shown in comparison to the state water quality standards for the receiving water.
- Grouped box and whisker plots for each monitored parameter were used to compare the influent and effluent concentrations. A Wilcoxon signed rank test was used to determine the significance.

Table 6. Summary of Bioretention Facility Pollutant Removal

	Mean Concentration		**Summation of Loads**			
Pollutant	**Average Inlet Concentration**	**Average Outlet Concentration**	**Sum of Inlet Load**	**Sum of Outlet Load**	**% of Pollutant Removed**	**State Water Quality Standards**
Nitrate-Nitrogen	3.5 mg/L	1.7 mg/L	7 kg	3.4 kg	51%	10 mg/L
Copper	66 µg/L	2 µg/L	131 g	4 g	97%	7 µg/L
Lead	42 µg/L	2 µg/L	83 g	4 g	95%	25 µg/L
Zinc	530 µg/L	25 µg/L	1 kg	950 g	95%	50 µg/L
Fecal Coliform	1111 / 100 mL	700 / 100 mL	2.2 x 1010	1.4 x 1010	37%	200 / 100 mL

*These hypothetical numbers are for illustration purposes only and are based on studies found in Winer (2000)
*The summation of loads approach assumes 70,000 ft^3/yr of stormwater influent.

Conclusion

The bioretention facility was found to reduce all pollutants to below the state water quality standards, with the exception of fecal coliform. However, only lead, copper, and zinc had sufficient removals to be statistically significant with adequate power. The other constituents would need additional sample pairs to make statistically relevant conclusions, although semi-quantitative results are possible. Metals had the highest pollutant removal, followed by nitrate-nitrogen, and fecal coliform. The monitoring team determined that in order to increase the effectiveness of fecal coliform removal, the design of the bioretention facility could be altered or used in conjunction with another stormwater treatment practice, as described in Study Design 4.

A comparison of inflow and outflow measurements determined that there was on average a 40% difference in inflow and outflow to the bioretention facility. This was found to be statistically significant at the 99% confidence level with a p-value < 0.01 using a Wilcoxon signed rank test. This suggests that widespread use of bioretention could be an effective element of a larger strategy to protect downstream channels from erosion.

Design

4

Implementation and Longevity Surveys of Stormwater Treatment Practices

What is the Study Design?

An increasing number of stormwater research studies seek to define the pollutant removal performance of individual stormwater treatment practices (STPs) so that communities can select the most appropriate devices to address their pollution concerns. Yet few communities evaluate how well these STPs actually function over time once they are installed, and in some communities there may be hundreds or even thousands of these practices. Implementation and longevity surveys of structural and non-structural STPs can determine whether an STP was installed to the design specifications, whether it is functioning as intended, and how well it holds up over time under its current maintenance regime. These surveys can provide critical data to improve STP design, refine local maintenance programs, and identify retrofit or repair opportunities.

Design standards for STPs are continuously evolving, and as a result, there is a lack of consistency in these standards across regions and time periods. This is particularly true for bioretention, infiltration, and more innovative types of STPs, such as low-impact development techniques. Gathering empirical data on STP function in the field can help to determine which design variants or specific design features are associated with the desired functions and, when paired with performance monitoring as outlined in Study Design 3, can link specific design features to enhance STP performance.

This Study Design involves a targeted field survey of a large population of STPs to evaluate performance indicators, maintenance and vegetative condition, longevity, aesthetics, or other factors related to their installation and intended functions. The survey primarily involves visual surveys but may also include limited hydrologic testing and, in some cases, interviews with adjacent residents to determine their attitudes about the STP. Visual screening is conducted for performance indicators, such as blockages, design or build failures, excessive sediment or pollutant accumulation, plant death, aesthetic appearance, and other factors. Figure 1 illustrates two bioretention facilities with the same design specifications (and presumably the same pollutant removal performance), one of which does not actually function properly due to installation or maintenance problems, which can be identified through a synoptic survey.

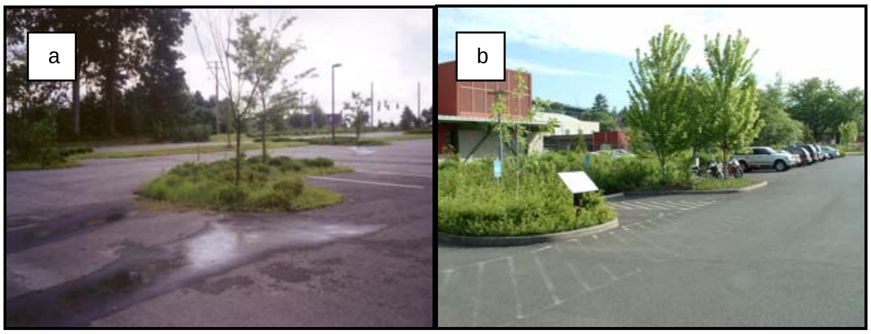


Figure 1. a) failing parking lot bioretention b) properly functioning parking lot bioretention

Although less than a dozen of this type of survey have been conducted, the results have been extremely important in identifying problems with existing STPs, as well as defining new directions in stormwater design and maintenance requirements. These synoptic surveys are relatively low cost, but can yield important stormwater management information that can be directly incorporated into local design manuals, development review procedures, and the day-to-day operations of the local stormwater program. Examples of these studies include:

- Investigation of pollution dynamics and habitat quality in stormwater ponds (Dewberry and Davis, 1990; Leersnyder, 1993; Oberts and Osgood, 1988; Bascietto and Adams, 1983; and Campbell, 1995).
- Assessment of failure rates and longevity of infiltration practices (Galli, 1993; and Hilding, 1993).
- Performance of biofilters and oil/grit separators (Reeves, 1995; and Shepp, 1995).
- Evaluation of how erosion and sediment controls are implemented at construction sites (Malcolm et al., 1990).

Why are Implementation and Longevity Surveys Needed?

The number of practical management issues that can be addressed through implementation and longevity surveys is almost limitless. The following examples illustrate the potential utility of STP surveys for the local stormwater manager.

- Evaluate which practices have a high failure rate and isolate the design or maintenance factors responsible for them (e.g., infiltration practices due to poor initial site selection, biofilters due to poor media selection).

- Determine the rate of sediment accumulation in pretreatment devices to develop an improved schedule for cleanout operations
- Systematically assess the mosquito breeding potential of different STP groups in the communities to develop better mosquito management data
- Evaluate the change in vegetative condition or cover within an STP group over time (e.g., constructed wetland, bioretention area)
- Compare actual geometry and design features of STPs with the recommended guidance
- Evaluate the performance and function of new or innovative STPs that have been recently installed in a community by supplementing the synoptic survey with collection of water quality samples at the inlet and outlet (see Study Design 3)
- Assess habitat quality in terms of desirable (e.g., native birds, amphibians) and undesirable species (e.g., geese, mosquitoes)
- Determine the thermal impact of the STP effluent to surface water through a specialized study that examines macroinvertebrates and vegetation coverage

What is the Basic Approach?

The basic approach is to conduct a survey of a large population of a single STP type in the community to evaluate specific factors related to its implementation, longevity and function. This field survey rigorously evaluates a pre-determined set of factors at each site during three separate sampling periods that include pre, post and during a storm event. The survey data is then compared to criteria and standards in local design manuals (e.g. adequate sizing to capture water quality volume, slope of swales), or other objectives (e.g. provide wildlife habitat, limit mosquito breeding, control trash) to determine whether the STP was implemented as designed and whether it is meeting its intended functions.

A key preparatory step is to develop a field form to collect the required data for the site investigation. The form should be designed to incorporate basic background information for the facility, such as age and maintenance responsibility, as well as detailed data pertaining to the monitoring objective(s). Attachment SD4-A to the Study Design provides example field forms to evaluate the performance for bioretention facilities (Characterization and Design – CAD) and Site Investigation and Assessment (SI). These are example survey forms and each community may use these as guidance to customize new forms based on individual needs.

What Factors should be Considered when Selecting Study Sites?

A systematic approach is recommended to select a representative population of STPs for the survey. This should begin with an investigation of local STP databases and/or original engineering plans and surveys to inventory the population of STPs available for study. Additional research may be needed to define the original design condition and purpose of the STP, as well as to collect some basic characteristics of its contributing drainage area. A review of past performance monitoring studies or maintenance records may also be helpful.

The next step is to identify the criteria that will be used to select the set of STPs to be surveyed. The criteria are based on the study design objective(s). Three general criteria to select the study sites include:

- Age – Age is a primary factor when selecting study sites as it often directly relates to the design standard on which construction was based. For example, if the objective of the study is to survey all ponds installed under the prevailing or current design manual, sites selected will all be of similar age and built since adoption of the current design standards. In some cases, older STPs may need to be surveyed in order to look at long-term issues related to maintenance needs, longevity, and/or vegetative successions.
- Representativeness – The selected STPs should be representative of conditions most commonly found within the jurisdiction (e.g., soil type, watershed location, and land use conditions which contribute to the varying pollutant sources and loads treated by the STPs).
- Accessibility – Good site accessibility is important for conducting the field investigation and ensuring the safety of the field crews.

After the selected criteria are applied to identify a set of STP sites for the survey, the STPs may be grouped, or stratified, by a key variable such as age of the facility, or percent impervious cover or predominant land use in the contributing drainage area. The purpose of stratification is to be able to compare results among the different groups of STPs to identify any important relationships. At least one site visit should be made to verify the site information for the final site selection.

What Unique Sampling Techniques and Equipment are needed?

The basic STP survey does not require much specialized sampling equipment and typically includes the material and equipment found in Table 1.

Field crews should have a basic understanding of STP functions and should be trained in the field survey methods, whether they are program staff, volunteers, or consultants. At least one member of the survey team should be a qualified design engineer who can verify the original STP design information and the field metrics collected. The crew leader should arrange for an orientation before going out in the field, possibly conducting a few example investigations so that the team members fully understand the process and how to correctly complete the field forms. For example, when a descriptive scoring/scale system is used on the field forms (e.g., scores range from 1 =poor condition to 5 =excellent condition), it is important for all members of the field crew to provide a consistent, objective evaluation.

Prior to the field investigation, it is recommended to organize field equipment, field forms, maps, etc. Sampling routes should also be planned out so that the field crews can be time efficient. During the site investigations it is important to thoroughly complete the field form, and also to take photos for further documentation of existing conditions and to provide a reference for possible future site investigations.

Table 1. Preparing for the Field Investigation

Basic Equipment	**Locator Map**
Clipboards and pencils Digital camera Scale and pocket calculator 100-foot measuring tape Pocket rod or local level Manhole puller Safety gear (cell phone, first aid kit, etc.)	Aerial photos Street names Hydrology Storm drain network STP locations Contributing drainage areas Property ownership
Materials	**Supplemental Equipment for Special Studies**
Field forms As-built engineering drawings Authorization letters Contact numbers for emergency assistance Photo IDs and business cards	*Sediment Quality* – Soil core sample, soil P test, soil toxicity test *Discharge Temperature* – hand-held temperature meter *Suspended Sediment* – depth integrated sampler *Turbidity* – turbidity probe *Litter* – floatable litter sampler

What Minimum Data is Needed to Characterize Drainage Area/Site Conditions?

To successfully undertake this type of monitoring, there is nearly as much preparatory work as there is fieldwork. The preparatory work is dependent upon a complete and accurate database of STPs; knowledge of current design standards, and if possible past design standards. It is recommended that an engineer, or someone with an engineering background, be a part of the monitoring team.

The specific parameters of interest for the survey will be driven entirely by the study objective and the type of STP in question. For example, if evaluating the mosquito breeding potential of stormwater wetlands, data such as ponding depth, water level fluctuations, and current maintenance practices would be important to document. Or, if the objective is to determine the rate of sediment accumulation in stormwater wet pond forebays, information on source areas within the contributing drainage area in addition to measuring sediment accumulation would be needed.

Once the parameters of interest are identified, a field form should be developed for use during the survey. The form should be designed to incorporate basic background information for the facility, such as age and maintenance responsibility, as well as detailed data pertaining to the monitoring objective(s) (see Attachment SD4-A). Basic information to be included in all survey forms prior to conducting the field work include:

- Contributing drainage area and boundaries,
- Characteristics of the drainage area (land use, land cover, percent impervious, drainage area, topography, soils),
- Precipitation and runoff volume captured estimates,
- Surveys of current topography,
- Design criteria,
- Characteristics of the STP (soils, vegetation, topography, depth to water table) as described above based on the study objective and STP type,
- Structural assessment of STP materials such as pipes, risers, outlets, and embankments,
- Maintenance practices that may be obtained by contacting the owner of the STP and/or its management company, and
- Any previous monitoring studies or surveys

Resources to acquire this information include: the original design or as-built drawing of the STP to obtain this data and characterize the site conditions, aerial imagery/photographs and any previous monitoring studies or surveys. Figure 2 provides an example of how referencing the as-built drawing of an STP can provide critical information to address the functioning or failure of a STP.

Once in the field, investigations will focus primarily on the STP, but may also be needed in the contributing drainage area to see if there have been changes since construction of the STP, or any noticeable source areas. A quick windshield survey of the drainage area can also help to identify drainage area factors that may influence STP performance (active construction sites, new development in off-site areas, etc). In addition, the area just downstream from the STP can be investigated to evaluate potential impacts from the STP.

Figure 2. While comparing an existing STP to the original plans, the field team realized the trash rack was never installed.

How much Sampling Effort is needed to get Reliable Data?

This study design is unlike other studies where an estimate for sample size can be made using statistical methods based on water quality variability. Less is known about the variability in STP installations, and the numerous design and maintenance factors that may affect performance capability; therefore, there is no strong statistical basis for estimating a minimum sample size for an implementation and longevity survey. Instead, as many STPs as possible should be investigated to generate a sufficient representation or characterization of the STP type. The general goal is to sample 25 STPs, with the realization that the size of the local STP inventory governs the available population that can be sampled.

Depending on the type of STP evaluated, it may be difficult to obtain a large sample size, as certain practices always seem to have a small population, such as infiltration and filtering practices (except for surface sand filters). This should not deter an evaluation as their investigation deserves merit, especially for new and innovative practices. Large populations are generally available for ponds, wetlands, and open channel practices.

In general, a minimum of three field surveys are needed for each STP to determine how the STP functions before, during and following a storm event. The timeframe for the survey immediately proceeding the storm is variable depending on the drain time of the STP (generally 48-72 hours). Surveying STPs before a storm is recommended to document dry conditions and to observe features of the pond that may be submerged during and following a rain event. The surveys should be timed to correspond with a period during the year that will maximize the opportunity to address the management issue (e.g. seasonal habitat for wildlife/breeding populations), or consider regional climate conditions to maximize survey efficiency. For example, trying to do a wet weather inspection may be more difficult during drier seasons as opposed to wetter seasons. Sampling opportunities are also restricted during snowmelt periods when a more limited number of dry STP conditions exist.

What are Special Data Management and Quality Control Considerations?

It is recommended to follow the basic elements of a QA/QC plan for the study design provided in Section 2 and Appendix G of this manual. In addition, the following are specific considerations. The main quality control issues with this specific study design include: ensure the collection of a consistent and robust dataset, and the ability to translate field observations into quantifiable and numeric data for analysis. The design of the field form should be developed with this in mind. Categories included on the survey forms should be well-defined and exclusive so that observations about STP function are as objective and non-ambiguous as possible. Field crews should be carefully trained to ensure consistency in data collection across different field teams.

Data collected from the survey should be analyzed to address the study objective. Categorizing and ranking methods and statistical tests can be used to make comparisons among the various sites investigated. Some suggested techniques are summarized below:

- Use rank order statistical tests to determine which design features most frequently lead to STP failure if the survey provides numerical results (e.g. Kruskall Wallace test), or tests for categorical data such as Chi-square. See Appendix B for additional information on statistical methods.
- Analyze data to determine what proportion of stormwater controls to answer the following question:
  - Require immediate maintenance?
  - Have invasive species? Algae growth?
  - Have downstream channel erosion? Erosion at inflow/outflow location?
  - Are fully treating runoff based on wet weather observations?
  - Are failing to meet stated design objectives?
- If the survey was conducted in conjunction with a performance monitoring study, the results can be used to identify specific design features that increase or decrease performance. An example is presented in Box 1.

Box 1. Design Point Method for Evaluating Design Features that Affect STP Pollutant Removal Performance

The design point method consists of a series of tables that award or deduct points for certain site-specific conditions and design factors present at the individual STP site. If the score is positive, the removal rate is higher than the accepted median pollutant removal for that specific practice. If the score is negative, the removal rate is lower than the accepted median pollutant removal for that specific practice. Refer to Schueler *et al.* (2007) for further details.

Bioretention Retrofits		
Design Factors	Points	
Exceeds target WQv by more than 50%	+ 3	
Exceeds target WQv by more than 25%	+ 2	
Tested filter media soil P Index less than 30 (phosphorus only)	+ 3	
Filter bed deeper than 30 inches	+ 1	
Two cell design with pretreatment	+ 1	
Permeable soils; no underdrain needed	+ 2	
Upflow pipe on underdrain	+1	
Impermeable soils; underdrain needed	- 1	
Filter bed less than 18 inches deep	- 1	
Single cell design	- 1	
Bioretention cell is less than 5% of CDA	-1	
Does not provide full water quality storage volume	- 2	
Filter media not tested for P Index (phosphorus only)	- 3	
NET DESIGN SCORE (max of 5 points)		
NET PHOSPHORUS SCORE (max of 5 points)		

How much should be budgeted for the Monitoring Study?

An example budget for this study design is provided in Table 3. This table provides general guidance only, and will vary with the local STP inventory, monitoring endpoints, site constraints, etc. The budget assumes that 25 STPs will be investigated during both wet and dry weather conditions.

Table 3. Budget for Monitoring the Performance of a Population of STPs

	Staff Resources	Unit Cost	Total Cost
Planning (17%)			
Background research (compile local STP inventory, secure GIS mapping layers)	40 hours	$50/hour	$2,000
Desktop analysis (major tasks include: preliminary site selection, preliminary site characterization, generate field maps)	32 hours	$50/hour	$1,600
Site visit to verify STP information prior to making the final site selection	32 hours	$50/hour	$1,600
Project Scope	16 hours	$50/hour	$800
Develop Monitoring Plan	8 hours	$50/hour	$400
Develop Field Forms	16 hours	$50/hour	$800
PLANNING SUBTOTAL			**$7,200**
Implementation (83%)			
Travel and Supplies			$2,000
Conducting the Study	4 hours/site investigation	$50/hour	$10,000[1]
Data Management (entering field data)	2 hours/site investigation	$50/hour	$5,000[1]
Data Evaluation	40 hours	$50/hour	$2,000
Final Report	100 hours	$50/hour	$5,000
IMPLEMENTATION SUBTOTAL			**$24,000**
TOTAL			**$31,200**

[1]Assumes 25 sites with 2 investigations per site (wet and dry weather conditions)

What Monitoring Problems can be anticipated?

Typical monitoring problems associated with this type of survey include:

- Inability to locate STPs listed in municipal databases or GIS maps,
- Incomplete or missing as-built drawing,
- Underground elements of the STP are difficult to corroborate with engineering plans, or
- Coordination of field survey times based on weather events can be challenging

What are some Good Monitoring Resources to Consult?

Galli, J. 1993. *Analysis of Urban BMP Performance and Longevity in Prince George's County, Maryland.* Metropolitan Washington Council of Governments, 202 pp.

Lampe, L., H. Andrews, M. Barrett, B. Woods-Ballard, R. Kellagher, P. Martin, C. Jeffries, and M. Hollon. 2005. *Performance and Whole Life Costs of Best Management Practices and Sustainable Urban Drainage Systems.* Project 01-CTS-21T. Water Environment Research Foundation.

Attachment SD4-A: Example Field Forms to Support Implementation and Longevity Surveys of STPs

Characterization and Design (CAD) is an example form for documentation of the site characteristics and design features of a bioretention facility. The form consists of a checklist to document such features as the contributing drainage area, facility size, type of pretreatment, conveyance system, etc.

The Site Investigation Assessment (SI) is also an example form for bioretention that scores the condition of the facility based on dry and wet weather survey conditions. Scoring criteria is based on the condition of such features as the inlets, pretreatment, facility, outlets, aesthetic and nuisance concerns, etc.

Bioretention
Assessment

Characterization and Design

CAD

FACILITY ID: ______	**WATERSHED/SUBSHED:**	**DATE:** ___/____/___	**ASSESSED BY:**
TIME:___:____AM/PM **LAT**___°_____'_____" **LONG** ___°_____'_____" **LMK:** _____ **MAPBOOK LOCATION:** __________			**GPS ID:**
OWNERSHIP ☐ Public ☐ Private	**YEAR CONSTRUCTED:** __________	**AS BUILT PLAN AVAILABLE:** ☐ Yes ☐ No *If yes, complete the remainder of this sheet prior to conducting the field survey.*	

SITE CHARACTERIZATION

CONTRIBUTING DRAINAGE AREA (%): *Note – All percentages should sum up to 100%.*

_____Industrial _____Commercial _____Urban/Residential _____Suburban/Res _____Forested _____Institutional
_____Golf course _____Park _____Crop _____Pasture _____Other: ___________

CONTRIBUTING DRAINAGE AREA MAINTENANCE: *Note – Complete during site investigation.*

1. Excessive trash/debris ☐ Yes ☐ No Comments: ______________________
2. Bare/exposed soil ☐ Yes ☐ No Comments: ______________________
3. Evidence of erosion ☐ Yes ☐ No Comments: ______________________
4. Excessive landscape waste / yard clippings ☐ Yes ☐ No Comments: ______________________

FACILITY SIZE	Surface Area: _____(ft^2)	**DRAINAGE AREA:** _____(acres)	**IMPERVIOUS COVER:** _____(%)
	Water Quality Volume: _____(ft^3)	**DESIGN STORM:** _____	

SOIL CORE SAMPLE: *Note – Complete during site investigation. Take within the bioretention facility.*
Dominant Soil Type ☐ Clay ☐ Loam ☐ Sand
Is the soil homogenous ☐ Yes ☐ No Comments: ______________________
Filter media soil P Index: __________

DESIGN ELEMENTS

HYDRAULIC CONFIGURATION
☐ On-line Facility ☐ Off-line Facility

TYPE OF PRETREATMENT FACILITY
☐ None ☐ Grass Filter Strip
☐ Sediment Forebay ☐ Plunge Pool
☐ Sediment Chamber ☐ Stone Diaphragm
☐ Grass Channel ☐ Other: __________

CONVEYANCE SYSTEMS

TYPE	**MATERIAL**
☐ Closed Pipe	☐ Concrete ☐ PVC/Plastic ☐ Metal ☐ Brick ☐ Other: __________
☐ Open Channel	☐ Concrete ☐ Earthen ☐ Other: __________

HYDROLOGY
Ponding Depth: __________(ft)
Depth to Water Table: __________(ft)

FEATURES (check all that apply)
☐ Outfall ☐ Inlet ☐ Outlet
☐ Riser ☐ Landscaping ☐ Maintenance Access
☐ Emergency Spillway
☐ Signage (if yes, Type:______________)

FIELD VERIFICATION NOTES

Identify any variations between the as-built and field observations.

Bioretention Site Investigation Assessment **SI**

FACILITY ID: ______	**WATERSHED/SUBSHED:**	**ASSESSED BY:**

MAINTENANCE AGREEMENT ☐ Yes ☐ No	**MAINTENANCE FREQUENCY** ☐ Weekly ☐ Monthly ☐ Annually ☐ Other: __________
MAINTENANCE PLAN ☐ Yes ☐ No	**DATE LAST MAINTAINED:** ___/___/___

PARTY RESPONSIBLE FOR MAINTENANCE: ________________________________
CONTACT: ________________________________
PHONE NUMBER: ________________________________
E-MAIL: ________________________________

DAY 1 – DRY WEATHER SURVEY **DATE:** ___/___/___

0 = Good condition. Well maintained, no action required.
1 = Moderate condition. Adequately maintained, routine maintenance needed.
2 = Degraded condition. Poorly maintained, routine maintenance and repair needed.
3 = Serious condition. Immediate need for repair or replacement.

PHOTO #'S:

RAIN IN LAST 72 HOURS ☐ Heavy rain ☐ Steady rain ☐ None ☐ Intermittent ☐ Trace

PRESENT CONDITIONS ☐ Clear ☐ Overcast ☐ Partly cloudy

INLETS

ITEM		**COMMENTS**
1. Elevation difference between pavement and inlet: ___________(ft)	0 1 2 3 N/A	
2. Depth of trash/debris accumulation: ___________(in)	0 1 2 3 N/A	
3. Depth of scour: ___________(in)	0 1 2 3 N/A	
4. Depth of sediment deposition: ___________(in)	0 1 2 3 N/A	
SCORE:		

PRETREATMENT

ITEM		**COMMENTS**
1. Maintenance access to pretreatment facility ☐ Yes ☐ No	0 1 2 3 N/A	
2. Depth of trash/debris accumulation: ___________(in)	0 1 2 3 N/A	
3. Depth of sediment deposition: ___________(in)	0 1 2 3 N/A	
4. Depth of scour/erosion: ___________(in)	0 1 2 3 N/A	
5. Depth of standing water: ___________(ft)	0 1 2 3 N/A	
6. Evidence of clogging ☐ Yes ☐ No	0 1 2 3 N/A	
7. % of pretreatment covered by dead vegetation/exposed soil: _____	0 1 2 3 N/A	
SCORE:		

FACILITY

ITEM		**COMMENTS**
1. Maintenance access to facility ☐ Yes ☐ No	0 1 2 3 N/A	
2. Depth of trash/debris accumulation: ___________(in)	0 1 2 3 N/A	
3. Depth of sediment deposition: ___________(in)	0 1 2 3 N/A	
4. Depth of scour/erosion: ___________(in)	0 1 2 3 N/A	
5. Depth of standing water: ___________(ft)	0 1 2 3 N/A	
6. Underdrain system ☐ broken ☐ clogged ☐ water ponding	0 1 2 3 N/A	
7. Depth of filter bed ponding: ___________(ft)	0 1 2 3 N/A	
8. Depth of mulch: ___________(in)	0 1 2 3 N/A	

9. Vegetation		
a. Plant composition consistent with as-built ☐ Yes ☐No	0 1 2 3 N/A	
b. % of facility covered by invasive species / weeds: _____	0 1 2 3 N/A	
c. % of facility covered by dead vegetation / exposed soil: _____	0 1 2 3 N/A	
SCORE:		

OUTLETS		
ITEM		COMMENTS
1. Depth of trash/debris accumulation: ___________(in)	0 1 2 3 N/A	
2. Depth of scour: ___________(in)	0 1 2 3 N/A	
3. Depth of sediment deposition: ___________(in)	0 1 2 3 N/A	
SCORE:		
MISCELLANEOUS		
ITEM		COMMENTS
1. Insect/mosquito problems ☐ Yes ☐No	0 1 2 3 N/A	
2. Number of animal burrows: _____	0 1 2 3 N/A	
3. Are safety measures (signage/fencing) present? ☐ Yes ☐No	0 1 2 3 N/A	
4. Presence of graffiti ☐ Yes ☐No	0 1 2 3 N/A	
5. Complaints from local residents ☐ Yes ☐No		
SCORE:		
TOTAL SCORE:		

DAY 2 – WET WEATHER SURVEY	DATE: ___/___/___
PHOTO #'S:	
PRESENT CONDITIONS ☐ Heavy rain ☐ Steady rain ☐ Intermittent ☐ Trace ☐ Overcast ☐ Partly cloudy	
RAIN GAUGE RESULTS Inches of rain from storm: ___________(in)	
FACILITY SKETCH	
Create a sketch of the facility showing conditions before and after the storm.	

Scoring (0-90) Total Score: _____

Good = 0 – 29; Moderate = 30 – 59; Degraded = 60 – 74; Serious = 75 – 90

Design

5

Monitoring Public Education Programs to Improve Water Quality

What is the Study Design?

Public education is a required element of MS4 stormwater management programs. It is the primary tool used to change attitudes, knowledge, and awareness of stormwater issues as well as to change behaviors, with the ultimate goal of improving water quality. Half of the six NPDES Phase II minimum management measures involve some form of public education. Therefore, it is important to design effective education programs that reach the targeted audience, reduce pollutant discharges, and protect water quality.

Despite the recognized utility of public education in managing stormwater pollutants, there is a lack of credible data to evaluate its effectiveness in terms of improving water quality. Most available studies provide data on the extent of the education effort that does not translate into meaningful, quantitative benchmarks to determine its effect on water quality. Very few studies have quantified the degree to which these programs attempt to change resident or business behavior, and even fewer have attempted to link behavior changes with improved stormwater quality (*Taylor et al.* 2007, Turner 2005, *Dietz et al.* 2004, Spetzman, no date).

This study design focuses on measuring changes in attitudes, awareness and behavior through surveys and other techniques as a result of public education efforts. It describes how to link behavior change to changes in water quality, which is measured either through outfall monitoring (described further in Study Design 1) or source area sampling. This study design includes several methods to adapt the source area sampling techniques described in Study Design 2 to measure water quality response associated with changes in behavior.

Education programs typically focus on promoting pollution-reducing behaviors known as source control practices (Figure 1). Source control practices can be implemented on residential, commercial, industrial, and institutional land to control nonpoint source pollution. In this study design, the example of public education programs geared toward residential areas and the general public is used. Schueler *et al.* (2004) describes the wide range of source control practices and implementation methods.

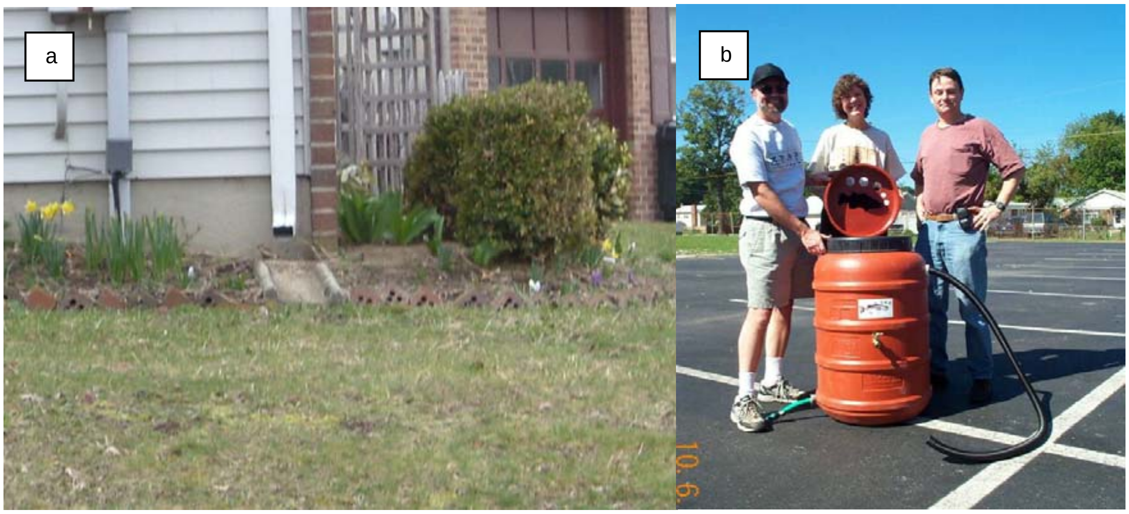


Figure 1. Elements of an education program to disconnect downspouts a) record number of downspouts connected/disconnected to storm drain system, b) workshop to build rain barrels.

Why is the Study Needed?

This study design can be used to:

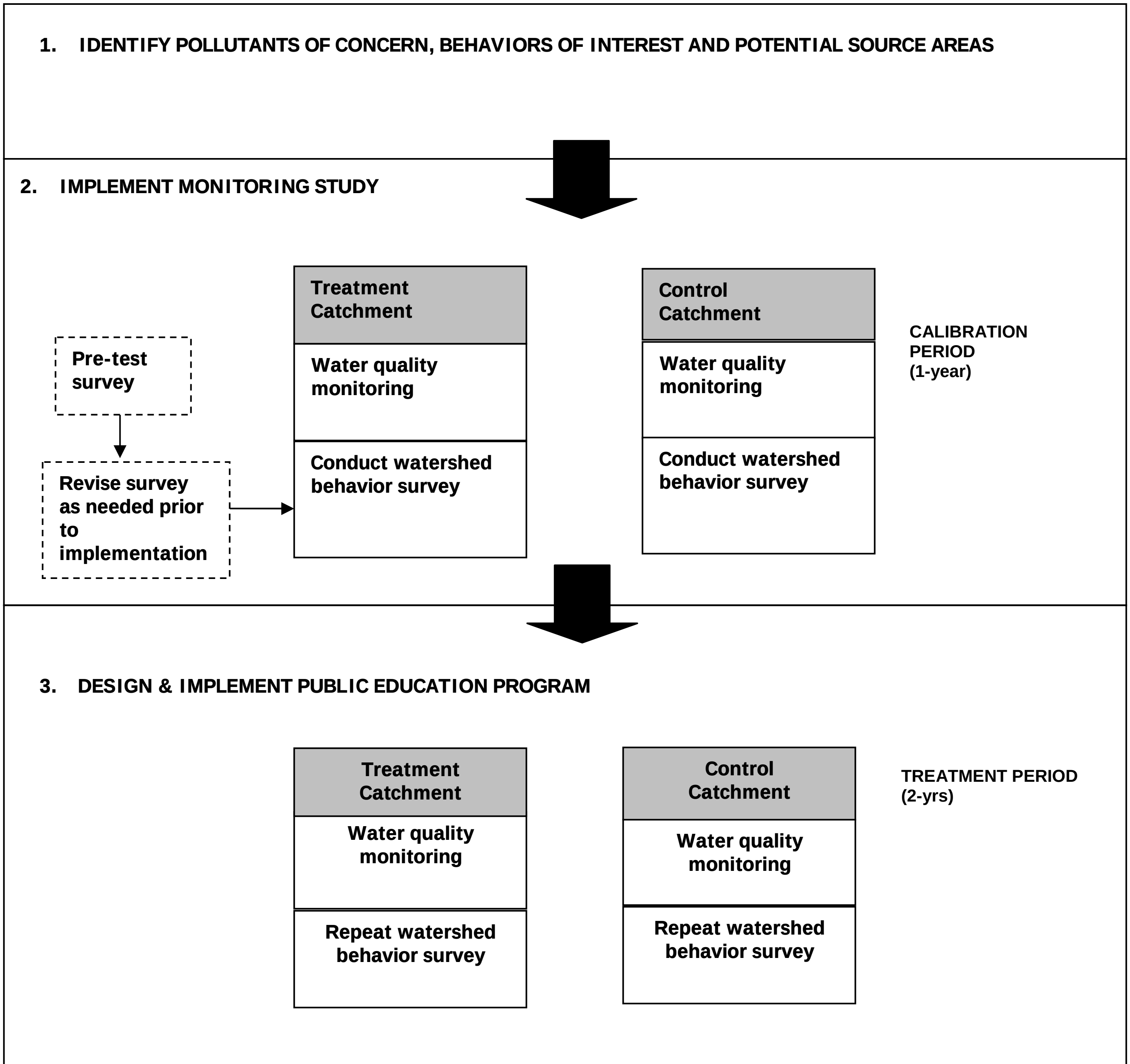
- Demonstrate the local water quality benefits associated with source control practices
- Determine which educational methods are most effective in getting the message out to the community
- Get reliable data on the most cost-effective investments in stormwater education and source control practices for the community
- Determine the combination of structural and non-structural methods that is most effective to reduce pollutant loadings, when done in conjunction with STP performance monitoring (described in Study Design 3).
- Define a defensible pollutant load reduction estimate as a result of the ongoing implementation of MS4 education programs

What is the Basic Approach?

The overall objectives of this study design are to: 1) measure the change in behaviors resulting from public education efforts and 2) determine if the change in behaviors affect a change in water quality. Both watershed behaviors and water quality are monitored throughout the study. The basic study design is a paired watershed approach, described in Study Design 6 which involves two study catchments, treatment and control, and calibration and treatment monitoring periods. Water quality can be measured either at the catchment outlet using outfall monitoring (see Study Design 1 for guidance) or at targeted pollutant source areas using source area sampling techniques (see Study Design 2 for guidance).

Figure 2 illustrates an example time frame of 3-years to coordinate the monitoring tasks associated with this study design. The actual time period for each study will vary based on the time needed to generate a significant relationship between the control and treatment catchments during the calibration period and enough samples to statistically evaluate the effect of treatment during the treatment period.

During the calibration period, water quality monitoring is conducted in both catchments to establish a statistical relationship between the study catchments prior to implementation of a treatment (which in this case is the public education program). A watershed behavior survey in both the treatment and control catchments is also conducted during the calibration period to establish a baseline of behavior. The education program is targeted towards a specific residential area with the goal of promoting the use of source control practice(s) to reduce the pollutant(s) of concern. *Schueler et al.* (2004) and *Wright et al.* (2004) provide guidance on identifying pollutants of concern, polluting behaviors of interest and potential source areas. Guidance to develop effective public education programs is provided in *Schueler et al.* (2004) and U.S. EPA (2003).



Resources:

- Pollutants of concern at an outfall: Study Design 1 in this manual
- Behaviors of interest: Schueler *et al.* 2004, Wright *et al.* 2004
- Potential source areas: Study Design 2 in this manual
- Paired watershed studies: Study Design 6 in this manual
- Public education programs: Schueler *et al.* 2004

Figure 2. Study design to monitor the effect of public education programs on water quality

The treatment period begins after the education program is implemented and source control practices have been in place for about one year to allow time to see a measurable result. For example, if the education program is intended to reduce the application of phosphorus-based fertilizers, the time period for monitoring should allow for at least one growing season to evaluate the impact post implementation. The same survey used during the calibration period is repeated in the treatment catchment while water quality monitoring continues in both catchments throughout the treatment period.

Water quality and survey data from the calibration and treatment periods for both the control and treatment catchments are compared to determine if the public education program has significantly affected water quality. Some statistical questions include:

- Is there a significant difference in watershed behavior between the calibration and treatment periods?
- Is there a significant difference in water quality between the control and treatment catchment during the treatment period?
- If there is a significant difference between the control and treatment catchments during the treatment period, can the difference in water quality be attributed to the education program?

Many references are available that describe the statistical analysis for paired watershed studies (e.g. Clausen and Spooner, 1993, Grabow et al. 1998).

What Factors Should I Consider when Selecting Study Sites?

The selection of study sites is critical to the success of the monitoring study. Well-chosen test sites make it easier to link the targeted pollutant with a watershed behavior and water quality, while poorly selected test sites can provide ambiguous results. Study Design 1 provides guidance on selecting test sites for outfall monitoring, while Study Design 2 provides guidance on selecting test sites for source area monitoring. In addition, the following factors should be considered when selecting study catchments for this study design:

- Catchments should be screened to determine where people most likely engage in the behavior that is directly linked to the pollutant of concern. Screening can be done using GIS or a field-based survey such as the Unified Subwatershed and Site Reconnaissance (Wright, *et al.* 2004). Selection of neighborhoods with the most severe pollution potential will more likely produce detectable downstream improvement.
- Population is likely amenable to behavior change/participation in program.
- Neighborhood generally has low turnover of residence ownership.
- Neighborhood drains to a perennial stream or STP that could be monitored for water quality (if outfall monitoring is part of the study design).

What Unique Sampling Techniques and Equipment Are Needed?

Sampling methods and equipment for outfall monitoring and source area sampling are provided in Study Design 1 and Study Design 2, respectively. Attachment SD5-A describes three methods to adapt source area sampling techniques to measure water quality response associated with behavior changes.

Behavior surveys are recommended to evaluate watershed behavior and its change over time. Survey types generally include: mail, telephone, in-person, internet (email or web-based), or some combination of the above. The type of survey should align with the purpose of the survey. Selecting the most appropriate type of survey depends upon the population (e.g. targeted or general public, geographically clustered or dispersed), the characteristics of the sample (e.g. homeowners, renters, those with internet access), types of questions to be asked (open or closed-ended), question topic (level of sensitivity), response rate desired, and available time and money. Table 1 presents some advantages and disadvantages of different types of surveys based on these factors.

Regardless of the survey type, survey designers should:

- Clearly articulate survey goal(s) to ensure the survey questions remain focused and are limited to the scope of the study,
- Use clear language that can be easily understood by respondents (typically Grade 5 comprehension level),
- Provide answers that are exclusive (e.g. multiple choice responses) and ensure the choice of responses will enable the respondent to provide an answer. This will help minimize interviewer bias
- Know how the survey data will be analyzed – e.g., what information the question will provide and what is the best way to represent that information,
- Format the survey for ease of completion and data entry, and
- Allow time to pre-test the survey.

The effectiveness of the study is in large part dependent upon how well the survey characterizes watershed behaviors and how the participants are selected for the survey. To keep the survey short and focused, a narrow list of watershed behaviors should be included. The Urban Subwatershed and Site Reconnaissance tool may be used to help narrow the types of behaviors to address in the survey (see Wright *et al.* 2004).

Table 1. Comparison of different types of surveys				
	Mail	Telephone	In-person	Internet
Relative total cost[1, 2, 3]	Moderate, $25-$35 per survey for 200 responses	Moderate $40 per interview May be as low as $12.50**	High $45-$60	Low * $2.50 - $5.50 for 2000 and 500 surveys respectively
Response time	Moderate-Long	Moderate	Long	Fast
Response rate	Low to moderate 5%-30%	Moderate 25% - 50%	High 40%-50%	Low to Moderate 2%-30%***
Staff requirements	Moderate	High	High	Low
Sample population limited****	No	Yes	No	Yes
Potential interviewer bias	No	Yes	Yes	No
Sampling need	Address	Phone number	Address	Email address
Complexity of questionnaire	Poor	Good	Best	Best
Control of sample+	No	Yes	Yes	No

[1] From Watt (1997)
[2] Owens (2002)
[3] Apostol and Irvine (no date)
* Email lowest cost but costs may significantly increase for web-based, custom programs
** No screening or special needs for a 5 minute telephone survey for 400 responses
***Post 2000 estimates, from 1986-2000 response rates ranged from 19%-61.5% (see Shaheen 2001)
**** Based on ability to generate random samples of population or limited access information such as unlisted phone numbers of households without computer internet access and email accounts.
+ the desired respondent is the same as the actual respondent

What Minimum Data are Needed to Characterize Drainage Area/Site Conditions?

The paired watershed approach requires the two catchments to be similar in physical characteristics, size and location. The following minimum data is needed to characterize the study catchments:

- Drainage area,
- Land use/land cover,
- Approximate age of development,
- Average lot size,
- Property owner contact information,
- Resident socio-economic conditions,
- Approximate neighborhood population,
- Storm drainage network (e.g., manholes, outfalls, STPs),

- Stream network, and
- Water and sewer service information.

If source area monitoring is used to measure water quality, additional data will be necessary to help select the best monitoring locations and quantify pollutant loadings. Table 2 lists the minimum survey data requirements to include in a survey for specific source areas. For quality assurance purposes, survey responses should be checked by the project lead or other designated individual to ensure consistency and completeness. Consult Study Designs 1 or 2 for additional information needed for outfall monitoring and source area monitoring, respectively.

Table 2. List of minimum survey data requirements for specific source areas

Lawns Amount of fertilizer applied Application frequency Type of fertilizer used (N-P-K ratio) Turfgrass area Runoff coefficient or a measure of soil compaction, density Clear area and yard slope for equipment placement	**Roof** Proportion of downspouts connected/disconnected to impervious surface Roof type and age Downspout material Precipitation chemistry
Streets Presence of curb and gutter Traffic volume Street sweeping practices Deicing practices Parking restrictions	**Parking Lots** Pavement conditions Other land cover types such as landscaped medians, Average parking density/or parking lot use such as % occupied Street sweeping practices

How Much Sampling Effort Is Needed to Get Reliable Data?

Two different sample estimates need to be determined as part of the study design, one for water quality and the other for watershed behavior. Paired samples are needed during the calibration and treatment periods. This includes simultaneously monitoring rainfall in both catchments in order to determine the variability of this most significant factor affecting runoff characteristics. Methods to determine the required sampling effort for water quality monitoring are provided in Appendix A.

A probability-based sampling method (e.g. random) is recommended to select the participants (sample) for the survey. This increases the likelihood that the data on watershed behaviors generated from the survey are representative of the population, and therefore the results may be used to infer watershed behaviors to the larger population. The sample should be taken from the same population for each of the two surveys (pre-treatment and post-treatment phase), and the participants for the survey (sample) can either be the same each time, or a new random sample selected. It is critical that the same survey be used each time. Appendix E describes various sampling methods for behavioral surveys and Box 1 provides some tips on conducting behavioral surveys.

To estimate the sample size for behavior surveys, the following factors should be considered:

- Project resources – how many surveys can be accommodated by staff or volunteers (e.g., for administering the survey and analyzing data)?
- Population size from which the sample is selected – is the population small or large and how does this affect sample selection and size? The smaller the population, the larger the proportion the sample size needed as there is greater variability or differences within smaller populations. Taylor (1998) recommends including the entire population in the sample if the population is 100 persons or less.
- Define the margin of error that provides a range of possible values in which the 'true' value lies of the population when used with the expected value.
- Level of confidence in results – it defines the risk of being wrong within the margin of error. A higher level of confidence will involve more samples than lower levels of confidence (e.g. 95% versus 80%). The lower the confidence level, the less certainty in the results (refer to Appendix E).
- The expected response rate of different survey types –a low response rate may yield inaccurate results.

Box 1. 10 Tips to Get Good Behavior Information (without spending a lot of money)

- Work on small neighborhood or catchment units that have a small population
- Use a group that the neighborhood already trusts (master gardeners, civic association or watershed group) to conduct the survey
- Train personnel conducting the survey
- Make it convenient
- Advance notification that affiliates the survey with a known organization or institution and provide identification cards of surveys that are in-person
- Personalize the request for information – be sure contact information is current and accurate (e.g. spelling, titles)
- Guarantee anonymity and confidentiality of information
- Provide some service or incentive (financial or non-monetary) to participate in the study
- Keep the survey short. Long surveys get less response than short survey (no more than ten) and have someone review the questions to check for any biases
- Ask questions that are within the comfort zone of the respondent.

How well the sample represents the population is based on the margin of error and confidence level. Most behavioral surveys define a 5-10% margin of error and a 95% confidence level. This may be difficult to achieve depending on the ability to generate a high response rate from surveys. For example, a study with a confidence level of 95% and a 5% margin of error that found 70% of homeowners fertilized their lawn may be interpreted as follows:

The results indicate with 95% confidence that 65% to 75% (e.g. 70% +/- 5%) of homeowners fertilize their lawns with a 5% chance of having a false positive result.

Therefore, knowing the population you are sampling from and selecting the most appropriate surveying tool to generate the highest response rate is most recommended (see Table 1).

There are many internet resources available to help estimate sample size using a pre-determined margin of error and confidence level (See Resources section). To get started, Appendix E provides a table and statistical equations to guide the selection of sample size for behavioral surveys. In general, the American Association for Statistics suggests to account for non-response rate of 30% and to increase the sample size accordingly (www.amstat.org). Applying the "10 tips to get good behavior information (without spending a lot of money)" should help to increase the response rate.

What are Special Data Management and Quality Control Considerations?

It is recommended to follow the basic elements of a QA/QC plan for the study design provided in Section 2 and Appendix G of this manual. In addition, the following are specific considerations for monitoring the impact of public education programs on water quality.

Surveys involve people and when information is requested, special care is needed to ensure the information is managed appropriately, interpreted objectively and reported accurately. Most types of surveys require data collected from survey sheets or forms to be input to another media such as spreadsheets or statistical software. Some tips for addressing the potential data management and quality control issues associated with surveys include:

- Consult municipal or department protocols to determine if there are specific procedures and permissions needed to collect and distribute information.
- Use a unique identifier for each survey to protect respondent anonymity and ensure confidentiality.
- The project coordinator should have sole access to personal information to protect anonymity and confidentiality.
- Use a well-designed and formatted survey to facilitate information transfer and input.
- Prior to data entry, coding of response categories may be needed (e.g. no = 0, yes = 1, don't know = 2). This may not be an issue for internet surveys that are designed and programmed to automatically compile the responses.
- For quality control purposes, a proportion of surveys (e.g. 15%) should be reviewed for accuracy of entries.
- Identify the survey non-responses requiring follow-up and track the response rate.
- Adopt a uniform procedure for making follow-up contact regarding missing or incomplete surveys (Scheurin 1998).

A variety of statistical techniques can be used to analyze the study results in order to answer the research questions addressed by the study. It should be noted that a 'lag effect' may exist where a response from the treatment may be longer than the monitoring period. For example, a comparison of means test may be used to determine if the watershed behaviors between the calibration and treatment catchment are statistically different. If there is a change in behavior and if no other changes occurred in either of these catchments during the treatment period, it may be determined with some level of certainty (as specified by of level of confidence) that the difference can be attributed to the education program, or treatment. Some statistical analyses to answer these questions include: a student's t-test for numerical, continuous data or Chi-Square test for categorical variables to determine the significant difference of survey responses between the calibration and treatment periods. An ANCOVA may be used to determine the significance of treatment on the reduction of pollutants for the calibration and treatment periods. A description of selected analytical methods is provided in Appendix B.

How Much Should I Budget for the Monitoring Study?

An example budget for a study to monitor the effectiveness of an education focused on downspout disconnection is provided in Table 3. Guidance on how to budget for outfall monitoring and source area monitoring is provided in Study Designs 1 and 2. These costs should be added to the total project budget for the control and treatment catchments. Detailed cost information for outreach techniques and water quality sampling and analysis is given in Appendix C.

The example budget shown in Table 3 is scoped for a three-year study to evaluate the impact of residential downspout disconnection on water quality. This is a pilot program to determine the effectiveness of an education program, targeted to a 10 square-mile residential catchment. A door-to-door survey of 200 randomly selected residents is administered annually, for a three-year period (one time during calibration period and once every year during the treatment period). The survey provided a financial incentive through a reduction in the municipal stormwater tax based on lot-level impervious cover. The incentive is used to achieve a minimum goal disconnecting 80% of the downspouts.

- In a GIS framework using overlay analysis, storm drain service area maps and parcel data are used to delineate catchment areas and identify the average age of development within each catchment.
- Targeted catchments included older single-family, detached homes built 1960-1980 in storm drain service areas. A Unified Subwatershed and Site Reconnaissance may be used to estimate the majority of downspouts that remain connected to the storm drain system (Wright et al. 2004). Control and treatment catchments are selected where the control catchment shared a similar proportion of connected downspouts and were similar in their age and type of development.
- During the calibration period, a field survey of the targeted catchments is used to estimate the proportion of homes with downspouts disconnected.
- Following the implementation of the education program in the treatment catchment, a behavior survey of 200 residents to estimate the proportion of homeowners who disconnected their downspouts. This is repeated in Year 3.
- Monitoring the control and treatment catchments at the outfalls over the three-year period.

Table 3. Example monitoring budget for a rooftop disconnection program		
	Staff Resources[1]	**Total Cost**
Planning (16%)		
Background research (data acquisition incl studies)	3 days	$1200
Desktop analysis (major tasks include: preliminary site selection, survey sample population, generate field maps)	7 days	$2,800
Project scope and sample design	3 days	$1,200
Develop monitoring plan	5 days	$2,000
SUBTOTAL		**$7,200**
Implementation (over 3-yr period) (84%)		
(note see Profile Sheet 1 for example source area monitoring budget)		
Supplies (GPS, cameras, street maps, postage* etc)		$1,000
Field Survey		
Perform USSR	16 staff days	$6,400
Survey		
Survey development	10 staff days	$4,000
Pilot survey[2,3]	25 hrs	$1,250
Revise survey as needed	1 day	$400
Implement survey[2] & follow-up	2 staff, 60 hrs each	$6,000
Training (both field and watershed behavior surveys)	2 staff, 24 hrs each	$2,400
Data Management		
Data QA/QC	16 hrs	$1,300
Data analysis and interpretation	10 days	$4,000
SUBTOTAL YEAR 1		$26,750
Repeat survey and source area monitoringYear 2[4]		$3,000
Repeat survey and source area monitoringYear 3[4]		$3,000
Final Report	5 days	$1,000
TOTAL		**$40,950**

[1] Assume $50/hr
[2] Allows 15 minutes per survey plus travel to site, cost will vary on survey method
[3] Administer 50 surveys, in person
[4] Cost of survey implementation

What Monitoring Problems Can Be Anticipated?

Study Designs 1 and 2 describe potential problems associated with source area monitoring and outfall monitoring. Other potential problems that may be encountered with conducting a behavioral survey include:

- Inherent difficulty in quantifying peoples' behaviors and attributing it to a public education program,
- Language barriers (demographics may require doing survey in different language),
- Difficulty reaching the age or demographic group associated with the behavior when it is small (e.g., do-it-yourself mechanics that dump oil in the storm drain),
- Inability to find a suitable control site,
- Low survey response rates,
- Insufficient funding to complete study or fund over a minimum of 3 years, or
- Inability to measure small water quality changes in response to behavior change given highly variable stormwater quality.

What are Some Good Monitoring Resources to Consult?

Public Education Programs

City of Austin Stillhouse Spring Cleaning project,
http://www.ci.austin.tx.us/growgreen/stillhouse.htm

New Hampshire's Scoop the Poop program,
http://www.des.state.nh.us/coastal/scoopthepoop.htm

New South Wales, Australia case studies on public education to address nonpoint source pollution, http://www.epa.nsw.gov.au/stormwater/casestudies/index.htm

Dietz, M. E., J. C. Clausen and K. K. Filchak. 2004. Education and changes in residential nonpoint source pollution. Environmental Management, 34(5): 684-690., http://www.joe.org/joe/2002december/rb5.shtml

Lake Harriet Watershed Awareness Project,
http://www.epa.gov/ORD/WebPubs/nctuw/Spetzman.pdf

Schueler, T. 2000. Understanding watershed behavior. The Practice of Watershed Protection. eds. T. Schueler and H. Holland. Center for Watershed Protection. Ellicott City, MD. pp. 621-628.

Schueler, T. 2000. On watershed education. The Practice of Watershed Protection. eds. T. Schueler and H. Holland. Center for Watershed Protection. Ellicott City, MD. pp. 629-635.

Survey Development

University of Wisconsin Cooperative Extension Program Development and Evaluation resources, http://www.uwex.edu/ces/pdande/evaluation/evaldocs.html

Designing Surveys and Questionaires, http://www.statpac.com/surveys/index.htm#toc

American Association of Statisticians (www.amstat.org)

www.whatisasurvey.info

Sample Size Estimate Calculators

http://www.raosoft.com/samplesize.html

http://www.surveysystem.com/sscalc.htm

http://survey.pearsonncs.com/sample-calc.htm

Paired watershed statistics

Clausen, J. C., and J. Spooner. 1993. *Paired Watershed Study Design*. 841-F-93-009. U.S. EPA Office of Water. Washington, DC.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1999. Detecting Water Quality Changes Before and After BMP implementation: Use of SAS for Statistical Analysis. In: NWQEP Notes, Number 93, January, 1999.

Other

Bacteria Source Tracking, http://www.epa.gov/OW-OWM.html/mtb/bacsortk.pdf

Pitt. R. 2007. Microorganisms in Urban Surface Waters. Available at: http://unix.eng.ua.edu/~rpitt/Class/ExperimentalDesignFieldSampling/MainEDFS.html

Attachment SD5-A

Adaptation of Source Area Monitoring Techniques to Link to Changes in Watershed Behavior

Example 1. Downspout Disconnection

- Conduct a behavior survey or field survey to determine the number of downspouts that are connected to the storm drain system in the study catchment both before and after treatment.
- Conduct source area monitoring of rooftops in the study catchment to determine average pollutant concentrations – or use numbers from existing source monitoring studies in your region (e.g. Bannerman et al. 2003, Steuer et al. 1997, Pitt and Voorhees 1995).
- Use existing local studies or conduct monitoring to characterize the chemical composition of precipitation. This will help to differentiate changes in pollutant loadings from downspout disconnection with the variability in precipitation.
- Use Geographic Information System (GIS) data or hard copy site plans to estimate the connected rooftop area both before and after treatment.
- Use the Simple Method (Schueler, 1987) or other pollutant load estimation method to estimate pollutant loads from rooftops in the study catchment both before and after treatment.

Table 4 provides example calculations for the total phosphorus load (TP) reduction that may be achieved using a public education program to encourage rooftop disconnection. The Simple Method (Schueler, 1987) is used to estimate pollutant loads.

Table 4. Example load estimate for total phosphorus from a downspout disconnection[1]						
CALIBRATION	**Source Area Monitoring**		**Behavior Survey Results**			
	Average TP (mg/L)[2]					
Residential roof	0.11		50% of homeowners downspouts disconnected			
Simple Method[3] to Estimate Phosphorus loading (lbs) before education program implemented						
Parameters	P	Pj	Rv	C	A	Load (lbs)
Residential roof	40	0.9	0.95	0.11	3.45	2.9
TREATMENT	**Source Area Monitoring**		**Behavior Survey Results**			
	Average TP (mg/L)					
Residential roof	0.11		75% of homeowners downspouts disconnected			
Simple Method to Estimate Phosphorus loading (lbs) after education program implemented						
Parameters	P	Pj	Rv	C	A	Load (lbs)
Residential roof	40	0.9	0.95	0.11	1.725	1.5

[1] Assumptions: 200 acre development, average housing footprint 1,000 ft^2, 300 houses
[2] Estimated from values reported in Steuer et al., 1997, Bannerman, 1993, and Waschbuch, 2000.
[3] Parameters for the Simple Method $L = [(P)(Pj)(Rv) \div (12)](C)(A)(2.72)$, where

L = Average annual pollutant load (pounds)
P = Average annual rainfall depth (inches)
P_j = Fraction of rainfall events that produce runoff
R_v = Runoff coefficient, which expresses the fraction of rainfall that is converted into runoff
C = Event mean concentration of the pollutant in urban runoff (mg/l)
A = Area of the contributing drainage (acres)
12 and 2.72 are unit conversion factors

Example 2. Reduced Fertilizer Education Program

- Conduct a behavior survey to estimate how much fertilizer is applied to lawns (and frequency) in the study catchment both before and after treatment.
- Use the results of behavior survey to identify lawns with varying levels of fertilizer application (none, moderate, high).
- Conduct source area monitoring to sample surface runoff from lawns and/or soil water leachate of each of type of lawn (e.g. none, moderate, high) using a sampler described by Waschbusch et al. 1999 or soil water samples such as lysimeters.
- Conduct soils or turfgrass analyses to quantify nutrient content (e.g., soil nitrate, soil P, turfgrass tissue P) from each type of lawn to supplement the source area sampling.
- Use results of behavior survey to estimate the amount of nutrients applied to lawns.
- Use the Simple Method (Schueler, 1987) or other pollutant load estimation technique to estimate before and after nutrient loads from lawns within the study catchment.

Box 2. How to estimate amount of fertilizer applied to lawns

Example Calculation:

It is common to express the amount of fertilizer applied in lbs per 1000 square feet (lbs/1000 ft²). This application rate will enable the amount of fertilizer applied to different lawn areas to be compared.

1. Determine the lawn area covered in turfgrass.
   Survey field work or other data provides the lawn area of
   . 10,000 ft²

2. The survey provides the following information:
   Fertilizer product information N-P-K ratio = 29-3-4
   (29% nitrogen, 3% phosphate, and 4% potash content)
   Amount of fertilizer used each application = one 40 lb bag
   Fertilizer frequency: 3 times per year = 120 lbs of fertilizer per
   . year

3. Determine the amount of nitrogen applied
   29% of 120 lbs is about 35 lbs of N.

4. Estimate the fertilizer application rate in lbs N per 1,000 ft2 per year.
   35 lbs of N is applied to the lawn area of 10,000 ft2
   The fertilizer application rate is estimated by dividing 35 by 10.

It is estimated that 3.5 lbs N per 1,000 ft2 is applied each year.

Go to City of Austin Stillhouse Spring Cleaning project for more information on a source area monitoring program to evaluate the impact of reduced fertilizer applications.

Example 3. Pet Waste Pickup

The method chosen to quantify pet owner's 'pick-up' practices considers the public and private locations of this activity (e.g. at home or at the park).

- Conduct a behavior survey to estimate how many dogs are in the study area, the proportion of dog owners that clean up after their pets, and the pickup frequency both before and after treatment. Direct observation over a several week period to determine pet waste pick-up behavior of pet owners is limited to public areas. Taylor-Powell and Steele (1996) describe how to use direction observation survey techniques.
- Collect surface runoff grab samples from lawns and public areas where survey data indicate the presence of dogs, or other household pets, or in-stream water quality samples using bacterial source tracking (BST). Risse and Jarrin (2004) use a targeted in-stream sampling approach that focuses on stream reaches with a high potential for bacterial contamination. Many BST techniques are experimental and can be expensive.
- Use the Simple Method (Schueler, 1987) to estimate before and after pollutant loads from lawns in the study catchment.

Go to New Hampshire's Scoop the Poop program for more information on how to set up a pet waste program.

Cumulative Effect of Treatment at the Catchment Scale

What is the Study Design?

Various approaches to stormwater treatment are being implemented by communities across the nation in order to mitigate the effects of land development on aquatic systems. These approaches have evolved over the years as practitioners, scientists and regulators learn more about the relationship between land development and aquatic systems and the performance and effectiveness of individual stormwater treatment practices (STPs). A broader question for which there is still a lack of data is how well a municipality's currently prescribed stormwater treatment regimen is working to actually change conditions in local streams. This study design helps to address the question by evaluating the cumulative effectiveness of stormwater treatment at the catchment scale (e.g., 50 to 500 acres).

Figure 1. Treatment Catchment in the Jordan Cove Watershed Study (graphic courtesy of University of Connecticut)

The term "treatment" is used throughout this paper to collectively refer to the combination of structural STPs and non-structural approaches implemented within the study catchment. Figure 1 illustrates a study catchment with treatment used in the Jordan Cove BMP Study in Connecticut.

Why is the Study Needed?

This monitoring protocol can be used to:

- Attribute changes in water quality, hydrologic, geomorphic, or biological conditions to a specific level of treatment.
- Determine if stormwater treatment is effective in meeting water quality, hydrologic, geomorphic or biological goals.
- Identify combinations of STPs that are most effective (for performance evaluation of individual STP, see Study Design 3 in this series).
- Make recommendations to change guidelines or criteria for stormwater treatment strategies.
- Provide measurable environmental results to meet municipal separate storm sewer system (MS4) stormwater program requirements.

What is the Basic Approach?

The basic study design is a paired watershed approach that tracks conditions in at least two similar catchments (control and treatment) during two study periods (calibration and treatment) (Clausen and Spooner, 1993). For a paired watershed study to be effective, it is important to choose catchments with similar physical characteristics (e.g., stream order, slope, level of impervious cover), size and location. This permits differences in stream conditions, such as biological diversity, flow, or pollutant loads, to be related to treatment assuming other conditions in the control and treatment catchments remain unchanged. The basic idea with a paired watershed approach is that a treatment (e.g., change in land use, management practices) will be implemented in one catchment, while a second catchment, the control, remains the same. Variations in the type of paired watershed studies are shown in Figure 2. The control catchment is used to account for year-to-year or seasonal climate variations. The paired watershed approach is also known as the before-after-control-impact or BACI approach referenced in other literature (Smith, 2002; Underwood, 1992).

Period	Catchment	
	Control	Treatment
Calibration	Undeveloped, no STPs	Undeveloped, no STPs
Treatment	Undeveloped, no STPs	Developed, STPs

a. Standard paired watershed study design

Period	Catchment		
	Control	Treatment1	Treatment2
Calibration	Undeveloped, no STPs	Undeveloped, no STPs	Undeveloped, no STPs
Treatment	Undeveloped, no STPs	Developed, STPs	Developed, STPs

b. Variation to examine multiple treatment catchments with different levels of treatment

Period	Catchment		
	Control1	Control2	Treatment
Calibration	Undeveloped, no STPs	Developed, no STPs	Undeveloped, STPs
Treatment	Undeveloped, no STPs	Developed, no STPs	Developed, STPs

c. Variation with both undeveloped and developed controls to separate out the affect of development from the effect of STPs.

Period	Catchment		
	Control1	Control2	Treatment
Calibration	Undeveloped, no STPs	Developed, no STPs	Developed, STPs
Treatment	Undeveloped, no STPs	Developed, no STPs	Developed, STPs retrofit

d. Variation where STPs are installed as a retrofit. In this case the 'treatment' is simply the retrofit of existing STPs as opposed to construction of new development with STPs.

Figure 2. Example Study Designs for Paired Watershed Studies

During the calibration period, a relationship between the control and treatment catchments is established by comparing conditions at their outlets using outfall monitoring techniques (described in Study Design 1). For example, if storm flow or water quality parameters are selected as indicators of condition, the collection of paired samples is required to determine how the two catchments respond to the same storm event. In this case, the relationship between control and treatment catchments is established by comparing flow, water quality or mass loadings for the paired samples. A pair of catchments may be considered sufficiently calibrated when an indicator for the control catchment can be used to predict the corresponding value for the treatment catchment (or vice versa) within an acceptable margin of error.

The calibration period must be long enough to obtain sufficient samples to statistically describe the relationship between study parameters in the treatment and control catchments. The exact time frame will vary depending on the desired number of samples and other factors such as, rainfall distribution, sampling frequency, and available resources. For example, a calibration period sampling runoff may take a full year to collect ten paired samples in the northeastern United States to establish a statistical relationship between the runoff in the control and treatment catchments. For other monitoring parameters that may be collected at a lower sampling frequency (e.g. indicators for macroinvertebrates) the calibration period may be longer. It is reasonable to budget one to two years for calibration for water quality parameters, but this may be shorter or longer depending on the precipitation patterns during this time period, monitoring problems that may be encountered, or variability in sampling parameters.

When the calibration period is complete, treatment is applied to the treatment catchment only and monitoring continues in both catchments for the remainder of the treatment period. While there is no recommended minimum time period for treatment monitoring, one can expect to spend about two years for the treatment period. The total timeframe for a paired watershed study may be as much as ten years or more to account for the time needed to construct the STPs being studied. Monitoring during the STP construction period is also valuable as it indicates the effectiveness of increasing levels of controls. However, in many cases, land disruption associated with the STP construction may more than negate any benefits from the additional controls. Therefore, some stabilization period may also be needed after that installation period. For controls relying on extensive vegetation, several growing seasons may be needed to obtain the full benefit of the control.

The effectiveness of treatment is determined by comparing conditions in the treatment catchment before and after treatment. If the difference in conditions between these two periods is statistically significant when compared with observed changes in the control catchment, the changes may be attributed to the treatment.

This type of study can be very complex, expensive, and time-intensive. However, the results are invaluable to determine whether stormwater treatment is effective in mitigating development impacts on streams. Partnerships with local universities, research institutions and federal agencies (e.g., USGS) are highly recommended for an MS4 wishing to conduct such a study. Coordination with local officials is essential to acquire the necessary permissions, approvals and exemptions from normal planning or development rules that may be needed to implement this type of project, and it is also important to coordinate closely with the local development community.

An alternative variation of the paired watershed study that is more economical but less rigorous, is to select a study catchment that has already been treated and substitute historical monitoring data to compare the effect of a treatment. This is referred to as a "time for space substitution" monitoring design. In this alternative study design, there is no calibration period, only a comparison between data collected prior to treatment and data from the treatment period. This design requires existence of sufficient pre-construction monitoring data for the study catchment selected, and may significantly reduce the study period. This reduction in the timeline and cost may allow for monitoring of more than one treatment catchment in order to compare the effectiveness of different types of treatment. Careful examination of the precipitation and flow data during the control (predevelopment) and treatment (post development) time periods is needed to ensure the data are comparable and not widely divergent (e.g. dry and wet years, hurricane activity). For example, the total rainfall, rainfall frequency and distribution between the two time periods can be compared to see if they are similar. This alternative approach is a time and money saver for a project and may be a reasonable substitution.

What Factors Should Be Considered when Selecting Study Catchments?

The site selection process is an intensive and challenging one given the variability among urban catchments in terms of land cover, drainage area, land use, and level of treatment. As no two catchments are truly identical and can also vary greatly from year to year, the calibration period eases the burden of finding 'identical' catchments by being able to statistically describe this variability. Generally, dozens of catchments must be screened to find acceptable treatment and control catchments. It is recommended to select sites that are proximal to one another to facilitate sample retrieval and maintenance of sites. Close proximity is also needed due to likely significant event to event variations in rainfall even over short distances in urban areas. Catchment selection criteria are presented in Table 1, which is followed by a description of three major challenges with catchment selection. The investigator is strongly encouraged to work closely with municipal staff to review land use data, development and stormwater management plans to facilitate the site selection process.

Table 1. Recommended Characteristics for Study Catchments

Catchment Type	**Recommended Characteristics**
All Study Catchments	▪ Control and treatment study catchments should have similar physical characteristics, such as size, slope, geology, and soils. ▪ Small size (e.g., 50 to 500 acres) ▪ Homogenous land use across the catchment ▪ Comprise the majority of the drainage area to a headwater stream (in order to minimize variability caused by factors other than treatment) ▪ No recent land use changes (for at least 2 years) ▪ Outlets have a stable channel and cross-section for discharge monitoring and do not leak at the outlet ▪ Pre-construction monitoring data is available (optional)
Treatment Catchment	▪ After treatment, will employ the combination of stormwater management strategies of interest ▪ Treatment will be applied across the entire catchment ▪ Other than the treatment of interest, no planned changes to land use or management practices during study period

Control Catchment	▪ No plans for land use changes or changes in management practices during the study period. ▪ Represents conditions in treatment catchment prior to development or implementation of treatment. ▪ Study may use two controls if desired to separate out the effect of development from the effect of STPs: 1. Undeveloped regional reference catchment that is predominantly forested or agricultural (whichever is most representative of pre-development land use in the area), 2. Developed catchment with no STPs that has been built out for at least 50 years (to ensure the stream has stabilized).

Timeframe

The first challenge with selecting catchments for the paired watershed approach is the relatively long (e.g., 4+ years) timeframe required for the study. This timeframe may make it difficult to select catchments in which changes in land use or management practices do not occur in the specified time period. Conversely, if the desired treatment is associated with new development, the time required to go from initial planning to final construction may be longer than the project funding allows. Working with municipal planners or other staff to identify parcels to be developed can help with catchment selection.

Representative level of treatment

The treatment catchment should be representative of the level or type of treatment for which the investigator wishes to evaluate effectiveness. For example, the Clarksburg Integrated Ecological Study in Maryland is evaluating the effect of newer "state-of-the-art" STPs as compared to older STPs (Jarnagin, no date). The treatment of interest in this case is construction of a conventional residential development with innovative STPs in series (dry wells, sand filters, infiltration areas, wet ponds). One of the control sites in this study is a developed residential catchment with conventional STPs (e.g., dry ponds, regional ponds). Finding catchments with the desired level of treatment may be difficult since an individual catchment usually contains a unique mix of STPs installed over several decades, designed under diverse design criteria, using different technologies and having various levels of functioning and maintenance. Another challenge is finding sites that are 100% treated or untreated.

Selection of control catchment

The control catchment should be representative of conditions in the treatment catchment prior to treatment. For example, in the Shepherd Creek paired watershed study in Cincinnati, Ohio, the treatment of interest is installation of parcel-level STPs (rain barrels and rain gardens) in a previously developed residential catchment (Roy et al. 2005). This study uses multiple control catchments: an undeveloped catchment that represents a reference condition, and a developed catchment with no STPs. The use of both undeveloped and developed controls helps to separate out the effect of development from the effect of STPs and can be used with most paired watershed studies. In most areas of the country, the reference or undeveloped condition will be forest or agriculture, but finding an undeveloped control catchment may be challenging in urban areas. Consultation with county or state-wide monitoring programs (e.g., Maryland Biological Stream Survey) can be helpful to identify local regional reference sites.

What Unique Sampling Techniques and Equipment Are Needed?

The sampling techniques used in a paired watershed study will vary depending on the selected monitoring parameters. Study Design 1 outlines the techniques and equipment required for water quality sampling at the outfall. Other possible sampling techniques (e.g., for hydrologic, physical or biological parameters) are listed in Table 2 with some resources for more information on the sampling protocol and required equipment.

Table 2. Monitoring Protocols for Paired Watershed Studies

Monitoring Parameter	**Methods/Data Source**
Stormflow monitoring	Richter et. al. (1996) USGS 2006
Stream channel cross sections	Stream Reference Sites (Harrelson et. al., 1994)
Geomorphic assessment	Ontario Ministry of the Environment (2003) Henshaw and Booth (2000) Richter et. al. (1996) Harrelson et. al.(1994)
Physical habitat	Rapid Bioassessment Protocol (Barbour et al. 1999) Rapid Stream Assessment Technique (Galli 1992)
Baseflow water quality sampling	Burton and Pitt (2002) Gordon et al. (2004) USGS 2006
Stormflow water quality sampling	Study Design 3 in this series Ely (2005) USGS National Field Manual for the Collection of Water Quality Data (USGS, no date)
Macroinvertebrate community	Rapid Bioassessment Protocol (Barbour et al. 1999) USGS National Water Quality Assessment Program methods (Cuffney et al., 1993)

What Minimum Data is Needed to Characterize Catchment Conditions?

A key element in any monitoring protocol is to carefully select the best combination of indicators to measure stream health, which in this study can be defined by water quality, geomorphic, habitat, biological, or hydrologic parameters. Selection of monitoring parameters will be driven by receiving water uses and the overall objectives of the study (i.e., they may be tied to specific watershed goals). It is recommended that multiple indicators be used, but not so many that the study becomes unaffordable and a data nightmare. As a general rule, monitoring parameters should be repeatable, sensitive, discrete, relatively inexpensive, and have a known distribution (Schueler and Kitchell, 2005). The range of monitoring parameters for a paired watershed study is wide, but may include those listed in Table 2. Additional monitoring data at selected STP locations may contribute useful information concerning the performance of the individual practices within this study design. Simultaneous monitoring at these different scales, while increasing the costs, will result in substantially more useful information to better infer the differences that may exist between the control and treatment catchments.

The minimum data needed to characterize conditions in the study catchments include:

- Drainage area,
- Soils, geology, slope,
- Land cover and land use,
- Existing pre-construction monitoring data including water quality, flow, stream geomorphology, habitat, and biological assessments (optional),
- Site plans and construction information for treatment catchment (if treatment will be implemented as part of a new development),
- Level and type of treatment proposed in treatment catchment (e.g., specific practices, quantity or quality goals, % of catchment treated, maintenance),
- Other obvious activities or impacts that may influence results,
- Precipitation data, and
- Various monitoring data based on indicators selected (e.g., flow, water quality, habitat).

How Much Sampling Effort Is Needed to Get Reliable Data?

- Sampling effort is essentially doubled in this study design where the goal is to collect paired samples – samples from both the control and treatment catchments for the same storm events. Paired samples are needed during the calibration and treatment periods. This includes simultaneously monitoring rainfall in both catchments in order to determine the variability of this most significant factor affecting runoff characteristics. Appendix A describes methods on how to estimate paired sample size.
- Parameter selection will determine the sampling frequency needed along with the sample size needed to detect a measureable change between the control and treatment catchments (e.g. stormflow monitoring, physical or biological).
- For studies that rely on behavior change as part of the treatment (e.g., education programs to reduce lawn fertilization), consult Study Design 5 for information on the sampling effort required to monitor the impact of public education efforts on stormwater quality.
- Look at the average number of runoff-generating storms on a regional basis as a starting point to estimate the possible number of storm sampling events. Consult Section 2 of this manual for more on rainfall analysis.

What Special Data Management and Quality Control Issues Can Be Expected?

- Potential bias is introduced into the study because sites are not randomly selected. The non-random selection should be acknowledged in study findings.
- Managing the chronology of paired rainfall, runoff volume and Event Mean Concentration (EMC) data for a range of storm events for the outfall is a central element of QA/QC
- A commonly used statistical method for a paired watershed study is an Analysis of Covariance (ANCOVA); however, a variety of other statistical analysis methods and software packages are available for use. Appendix B provides information on statistical tests that may be used for paired watershed studies.

How Much Should Be Budgeted for the Monitoring Study?

A paired watershed study is likely to cost a minimum of $250,000 and up to $500,000 or more. For example, the Jordan Cove Urban Watershed Project cost $55,000 to manage the project and $1.3 million for the monitoring (Clausen, 2007). This project used a paired watershed study design to evaluate the effectiveness of several STPs on water quality in a newly developed catchment in Connecticut. The monitoring portion was very expensive because it focused on water quality sampling and took place over a 10-year period. The importance of partnering with federal and academic partners cannot be stressed enough. These partners can offset the cost to the MS4 by providing staff time or by utilizing existing monitoring stations and resources as part of the paired watershed study.

An example budget for a paired watershed monitoring study is provided in Table 3. This table provides general guidance that can be adapted based on parameters such as number of samples or length of study period. Project costs will vary with the monitoring parameters selected and variability of data. Some assumptions for the Table 3 budget include:

- Budget does not include costs associated the purchase or installation of treatment practices.
- Calibration period is two years and treatment period is two years.
- Parameters of interest include:
  - Intensive series of cross-sections measured biannually (10 representative cross-sections in each catchment),
  - Continuous flow monitoring,
  - Turbidity – continuous measurement, and
  - Stormflow sampling for total suspended solids (6 storm events per year).
- Laboratory Analysis includes analysis of storm samples for TSS (average 10 samples taken over the course of each storm event). Cost = $15 per sample plus shipping and handling.

- Storm sample collection requires two staff and two hours per storm.
- Cross sections require two staff and two hours each.
- Equipment maintenance in required two times per month (two staff, two hours each) in each catchment.

Table 3. Budget for Monitoring the Cumulative Treatment Effect			
	Staff Resources	Unit Cost	Total Cost
Planning (20%)			
Background research (determine the control and treatment catchments)	40 hours	$50/hour	$2,000
Desktop analysis (site characterization, generate field maps, determine cross-section locations)	40 hours	$50/hour	$2,000
Project Scoping	32 hours	$50/hour	$1,600
Develop Monitoring Plan	32 hours	$50/hour	$1,600
Project Management	200 hours/yr	$50/hour	$50,000
PLANNING SUBTOTAL			**$57,200**
Implementation (80%)			
ISCO sampler with flow meter (2)		$10,000	$20,000
YSI6000 Turbidity optical sensor (2)		$5,000	$10,000
Sokkia Total survey Station (1)		$6,000	$6,000
Digital camera (1)		$200	$200
Equipment Installation	64 hours	$50/hour	$3,200
Calibration Monitoring (2 years)	400 hours/yr	$50/hour	$40,000
Treatment Monitoring (2 years)	400 hours/yr	$50/hour	$40,000
Laboratory Analysis (for 10 storm events per year)		$1,500/yr	$7,500
Data Management	100 hours/yr	$50/hour	$25,000
Data Evaluation	200 hours/yr	$50/hour	$50,000
Final Report	250 hours	$50/hour	$12,500
IMPLEMENTATION SUBTOTAL			**$201,600**
TOTAL			**$258,800**

What Monitoring Problems Can Be Anticipated?

- Difficulty finding study catchments with similar characteristics, minimal variability in land cover, and desired level of treatment
- Longer timeframe may be needed to detect changes from treatment
- Development in the treatment catchment may not occur at the desired speed due to development or planning issues
- Frequent equipment maintenance and failure
- Vandalism and safety concerns in urban catchments
- Maintaining continuity of staff and consultants over the long timeframe

- Funding runs out before sufficient sampling data is obtained
- The control site is developed or management practices change before the study is complete

What are Some Good Monitoring Resources to Consult?

Barr Engineering Company. 2006. *Burnsville Stormwater Retrofit Study*. Prepared for the City of Burnsville, MN.

Clausen, J. C. 2007. *Jordan Cove Watershed Project Final Report.* University of Connecticut. Storrs, CT.

Clausen, J. C., and J. Spooner. 1993. *Paired Watershed Study Design.* 841-F-93-009. U.S. EPA Office of Water. Washington, DC.

Cheng, M., Coffman, L.S., Zhang, Y., and Z. J. Licsko. 2005. Hydrologic responses from low impact development compared to conventional development.

Coombes P., J. Argue and G. Kuczera. 2000. Figtree Place: A case study in water sensitive urban development. *Urban Water Journal* 4(1). London, UK.

Dietz, M. E., and J. C. Clausen. 2007. Stormwater Runoff and Export Changes with Development in a Traditional and Low Impact Subdivision. *Journal of Environmental Management*, in press.

Grabow, G., Spooner, J., Lombardo, L., Line, D. E., and K. L. Tweedy. 1998. *Has Water Quality Improved?: Use of a Spreadsheet for Statistical Analysis of Paired Watershed, Upstream/Downstream and Before/ After Monitoring Designs.* NCSU Water Quality Group, Biological and Agricultural Engineering Department, NC State University. Prepared for the Sixth National Nonpoint-Source Monitoring Workshop, September 21-24, 1998. Cedar Rapids, IA.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1999. Detecting Water Quality Changes Before and After BMP implementation: Use of SAS for Statistical Analysis. In: NWQEP Notes, Number 93, January, 1999.

Helsel, D.R. and R.M. Hirsch. Statistical Methods in Water Resources. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Data and Interpretation

Jarnagin, S. T. No date. *Collaborative Research: Streamflow, Urban Riparian Zones, BMPs and Impervious Surfaces.* Online: http://www.epa.gov/esd/land-sci/epic/pdf/usepa_clarksburg_research_nov2006.pdf (Accessed February 20, 2008).

Lloyd, S., T. Wong and C. Chesterfield. 2002. *Water sensitive urban design: a stormwater management perspective*. Cooperative Research Centre for Catchment. Industry Report 02/10Monash University, Victoria 3800 Australia.

Roy, A. H., Mayer, A. L., Shuster, W. D., Thurston, H. W., Hoagland, T., Clagett, M. P., Parikh, P. K., and M. A. Taylor. 2005. *A multidisciplinary approach to stormwater management at the catchment scale.* 10^{th} International Conference on Urban Drainage. Copenhagen/Denmark, 21-26 August 2005.

Schueler, T.R., Sturm, P. E., and M. Paul. No date. Research Protocol to Evaluate the Effectiveness of BMPs and Site Treatments to Mitigate the Impacts of Development on Streams. Prepared for the National Association of Home Builders. Center for Watershed Protection. Ellicott City, MD.

U.S. Department of Agriculture, Natural Resources Conservation Service. 1996. *National Handbook of Water Quality Monitoring*. Part 600 National Water Quality Handbook.

University of Connecticut. No date. Jordan Cove Urban Watershed Project. University of Connecticut College of Agriculture and Natural Resources. http://www.cag.uconn.edu/nrme/jordancove/

U.S. Geological Survey Eastern Geographic Science Center. No date. *Clarksburg Integrated Ecological Study Workshop Highlights.* http://erg.usgs.gov/clarksburghighlights.html (Accessed February 20, 2008).

U.S. Geological Survey Eastern Geographic Science Center. 2006. National Field Manual for the Collection of Water-Quality Data. Chapter A4. Collection of Water Sample. Techniques of Water-Resources Investigations Book 9. http://water.usgs.gov/owq/FieldManual/chapter4/pdf/Chap4_v2.pdf

Section 4

Resources for More Information

Databases

National Stormwater Quality Database, Version 3. Robert Pitt, University of Alabama. http://unix.eng.ua.edu/~rpitt/Research/ms4/mainms4.shtml

International Stormwater (NSW) Best Management Practices (BMP) Database. 2006. http://www.bmpdatabase.org

Urban BMP Performance Tool. U.S. EPA National Pollutant Discharge Elimination System. http://www.epa.gov/npdes/urbanbmp

National Pollutant Removal Performance Database, Version 3. Center for Watershed Protection. http://www.cwp.org

National Oceanic and Atmospheric Administration, National Weather Service Forecast Office Rainfall records. http://www.srh.noaa.gov/ffc/html/rainresrc.shtml

Sample collection techniques

Field protocols for field preparations and preventing sampling contamination National Field Manual for the Collection of Water Quality-Data. Chapter A4 Collection of Water Samples. U.S. Geological Survey, Techniques of Water-Resources Investigations. http://water.usgs.gov/owq/FieldManual/chapter4/pdf/Chap4_v2.pdf

Stormwater Effects Handbook: A Toolbox for Watershed Managers, Scientists and Engineers. Allen Burton and Robert Pitt, 2000 http://unix.eng.ua.edu/~rpitt/Publications/Publications.shtml

Example water quality sample field sheet may be downloaded at:

Water Quality Sampling Field Data Sheet. U.S. EPA http://www.epa.gov/owow/monitoring/volunteer/stream/ds5b.pdf

Selected references on sampling gross pollutants in stormwater runoff

ASCE Guideline for Monitoring Stormwater Gross Pollutants, Draft. Environmental Resources Institute, Urban Water Resources Research Council, and Gross Solids Technical Committee.
http://www.stormwaterauthority.org/library/library.aspx?id=188

Breault, R.F., K. P. Smith and J. R. Sorenson. 2005. Residential street-dirt accumulation rates and chemical composition, and removal efficiencies by mechanical- and vacuum-type sweepers, New Bedford, Massachusetts, 2003-04. USGS Scientific Investigations Report 2005-5184. U.S. Department of the Interior.

QA/QC

Numerous U.S. EPA general guidance reports on preparing QA/QC plans:
http://www.epa.gov/quality/qa_docs.html#guidance

Volunteers

EPA's national directory of volunteer monitoring programs can be used to search for a local program:

National Directory of Volunteer Monitoring Programs, U.S. EPA.
http://yosemite.epa.gov/water/volmon.nsf/Home?readform.

Volunteer Stream Monitoring: A Methods Manual. U.S. EPA Office of Water.
http://www.epa.gov/volunteer/stream/stream.pdf

Volunteer Monitoring, U.S. EPA. http://www.epa.gov/owow/monitoring/volunteer/

Land Grant Colleges' and Universities' Volunteer Water Quality Monitoring National Facilitation Project. http://www.uwex.edu/ces/csreesvolmon/

Save Our Streams program, The Izaak Walton League of America.
http://www.iwla.org/index.php?id=19

NPDES

Overview of the U.S. EPA Stormwater Program. U.S. EPA.
http://www.epa.gov/npdes/stormwater

Measurable Goals Guidance for Phase II Small MS4s. U.S. EPA National Pollutant Discharge Elimination System.
http://www.epa.gov/npdes/pubs/measurablegoals.pdf

Stormwater Phase II Final Rule Fact Sheet Series. U.S. EPA National Pollutant Discharge Elimination System. http://cfpub.epa.gov/npdes/stormwater/swfinal.cfm

References

American Society of Civil Engineers (ASCE) and United States Environmental Protection Agency (EPA). 1999. *Development of Performance Measures. Task 3.1 – Technical Memorandum. Determining Urban Stormwater Best Management Practice (BMP) Removal Efficiencies*. Available at: http://www.bmpdatabase.org/Docs/task3_1.pdf

American Society of Civil Engineers (ASCE) and United States Environmental Protection Agency (EPA). 2002. *Urban Stormwater BMP Performance Monitoring; a guidance manual for meeting the national stormwater BMP database requirements*. EPA-821-B-02-001

Apostol, T. and R. Irvine. No date. *Research Fundamentals: Estimating Telephone Survey Research Costs*. WestGroup Research http://www.westgroupresearch.com/research/phonecosts.html (accessed July 2007).

ASTM (American Society of Testing and Materials). 1996 *Annual Book of ASTM Standards*. West Conshohocken, PA: ASTM, vol. 04.08.

Bannerman, R., D. Owens and N. Hornewer. 1993. "Sources of Pollutants in Wisconsin Stormwater." *Water Science Technology,* 28(3-5): 241-259.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition.* EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Bascietto, J., and L.Adams. 1983. "Frogs and Toads of Stormwater Management Basins in Columbia, MD." *Bulletin of MD Herpetological Society* 19(2): 58-60.

Bochis, C., R. Pitt, and P. Johnson. 2008. Land development characteristics in Jefferson County, Alabama. In: *Stormwater and Urban Water Systems Modeling*, Monograph 16. (edited by W. James, E.A. McBean, R.E. Pitt and S.J. Wright). CHI. Guelph, Ontario, pp. 249 – 282.

Brown, E., D. Caraco and R. Pitt. 2004. *Illicit Discharge Detection and Elimination: a guidance manual for program development and technical assessments*. Center for Watershed Protection and University of Alabama. EPA X-82907801-0.U.S. EPA Office of Wastewater Management, Washington, D.C.

Bucheli, T. D., S. R. Muller, S. Heberle, and R. P. Schwarzenback. 1998. "Occurrence and behavior of pesticides in rainwater, roof runoff and artificial stormwater infiltration." *Environmental Science and Technology*, 32: 3457-3464.

Burton, A., and R. Pitt. 2002. *Stormwater Effects Handbook: a Toolbox for Watershed Managers, Scientists, and Engineers*. Lewis Publishers. New York, NY.

Campbell, K.R. 1995. "Concentrations of Heavy Metals Associated with Urban Runoff in Fish Living in Stormwater Ponds." *Archives of Environmental Contamination and Toxicology,* 27: 352-356.

Center for Watershed Protection. 1998. *Rapid Watershed Planning Handbook: A Comprehensive Guide for Managing Urbanizing Watersheds*. Ellicott City, MD.

Center for Watershed Protection. 2005. *Final Report. Stormwater Management Strategy for Catchment O in Watershed 263*. Parks and People Foundation and Cheseapeake Bay Trust. Ellicott City, MD.

Center for Watershed Protection. 2006. *Monitoring Plan. Deriving Reliable Pollutant Removal Rates for Municipal Street Sweeping and Storm Drain Cleanout Programs in the Chesapeake Bay Basin*. Ellicott City, MD.

Center for Watershed Protection. 2006. Street Sweeping Monitoring Plan.

Gray, J.R., G. D. Glysson, L. M .Turcios and G. E. Schwarz. 2000. *Comparability of Suspended-Sediment Concentration and Total Suspended Solids Data*. U.S. Geological Survey, Water-Resources Investigations Report 00-4191. Reston,VA.

Center for Watershed Protection. 2007. *National Pollutant Removal Performance Database for Stormwater Treatment Practices*. Second edition. Center for Watershed Protection, Ellicott City, MD.

Center for Watershed Protection. 2008. *BMP Evaluation Tool*. Center for Watershed Protection, Ellicott City, MD.

Chang, M., M. W. McBroom and R. S. Beasley. 2004. "Roofing as a source of nonpoint source pollution." *Journal of Environmental Management,* 73:307-315.

Clark, S.E. and R. Pitt. 2008. Comparison of stormwater solids analytical methods for performance evaluation of manufactured treatment devices. *Journal of Environmental Engineering*, Vol. 134, no. 4, 259-264.

Clausen, J. C., and J. Spooner. 1993. *Paired Watershed Study Design*. 841-F-93-009. U.S. EPA Office of Water. Washington, DC.

Clausen, J. C. 2007. *Jordan Cove Watershed Project Final Report*. University of Connecticut. Storrs, CT.

Council of State Governments. 1998. *Getting in Step: A Guide to Effective Outreach in Your Community*. Lexington, KY.

Cuffney, T.F., Zappia, H., Giddings, E. M. P., and Coles, J. F. 2005. Effects of urbanization on benthic macroinvertebrate assemblages in contrasting environmental settings: Boston, Massachusetts, Birmingham, Alabama, and Salt Lake City, Utah. Proceedings of Symposium 47: Effects of urbanization on stream ecosystems, American Fisheries Society, Bethesda, MD, 361-407.

Davis, A.P. 2006. Personal Communication. University of Maryland College Park.

Dengler, J. and J. Brasino. 2007. "A practical measure of the impact of driveway car washing." *Stormwater Magazine*, October 2007. Available online at http://www.gradingandexcavating.com

Dewberry and Davis, Inc. 1990. *Toxicity of Sediments from BMP Ponds*. Final Report. Northern Virginia Planning District Commission. Annandale, VA.

Dietz, M. E., J. C. Clausen and K. K. Filchak. 2004. "Education and changes in residential nonpoint source pollution." *Environmental Management*, 34(5): 684-690.

Ely, E., 2005. "Russian River First Flush Volunteers Capture Storm Runoff." *Lakeline,* 25 (1).

Galli, J. 1992. *Rapid Stream Assessment Technique*. Metropolitan Washington Council of Governments. Washington, DC.

Garn, H.S. 2002. *Effects of Lawn Fertilizer on Nutrient Concentration in Runoff from Lakeshore Lawns, Lauderdale Lakes, Wisconsin*. USGS Water-Resources Investigations Report 02-4120.

Glass, C. 2006. Personal Communication. Howard University.

Gordon, N. D., McMahon, T. A., Finlayson, B. L., Gippel, C. J., and R. J. Nathan. 2004. *Stream Hydrology: An Introduction for Ecologists*. John Wiley and Sons Ltd. West Sussex, England.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1998. "Detecting Water Quality Changes Before and After BMP implementation: Use of a Spreadsheet for Statistical Analysis." *NWQEP Notes*, Number 92, November, 1998. Available at http://www.bae.ncsu.edu/programs/extension/wqg/issues/pubindex.html (accessed February 2008)

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1999. "Detecting Water Quality Changes Before and After BMP implementation: Use of SAS for Statistical Analysis." *NWQEP Notes*, Number 93, January, 1999. Available at http://www.bae.ncsu.edu/programs/extension/wqg/issues/pubindex.html (accessed February 2008)

Gray, J.R., G. D. Glysson, L. M .Turcios and G. E. Schwarz. 2000. *Comparability of Suspended-Sediment Concentration and Total Suspended Solids Data*. U.S. Geological Survey, Water-Resources Investigations Report 00-4191. Reston,VA.

Harrelson, C.C., Rawlins, C. L. and J.P. Potyondy. 1994. *Stream Channel Reference Sites: an Illustrated Guide to Field Techniques*. Gen. Tech. Rep. RM-245. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

Helsel, D. R. 2005. *Nondetects and Data Analysis – Statistics for Censored Environmental Data*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Henshaw, P.C., and D.B. Booth. 2000. "Natural restabilization of stream channels in urban watersheds." *Journal of the American Water Resources Association,* 36:1219-1236.

Hilding, K. 1993. *A Study of Infiltration Basins in the Puget Sound Region.* M.E. Thesis. Dept. of Biological and Agricultural Engineering. Univ. of California, Davis.

Horowitz, A.J., T. S. Hayes, J.R. Gray and P.D. Capel. 1997. *Selected laboratory tests of the whole-water sample splitting capabilities of the 14-liter churn and Teflon cone splitters.* U.S. Geological Survey Office of Water-Quality Technical Memo 97-06. 24 pp.

International Stormwater (NSW) Best Management Practices (BMP) Database. 2006. Available online at: http://www.bmpdatabase.org

U.S. EPA's Urban BMP performance evaluation tool Available online at:http://cfpub.epa.gov/npdes/stormwater/urbanbmp/bmpeffectiveness.cfm

Jarnagin, S. T. No date. *Collaborative Research: Streamflow, Urban Riparian Zones, BMPs and Impervious Surfaces.* Online: http://www.epa.gov/esd/land-sci/epic/pdf/usepa_clarksburg_research_nov2006.pdf (Accessed February 20, 2008).

KCI Technologies Inc. 2004. *Watershed 263 Management Plan*. Prepared for the City of Baltimore Department of Pubic Works, Water Quality Management Office. Hunt Valley, MD.

Kitchell, A. and T. Schueler. 2004. *Unified Stream Assessment: A User's Manual.* Manual 10 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection. Ellicott City, MD.

Lampe, L., H. Andrews, M. Barrett, B. Woods-Ballard, R. Kellagher, P. Martin, C. Jeffries, and M. Hollon. 2005. *Performance and Whole Life Costs of Best Management Practices and Sustainable Urban Drainage Systems.* Project 01-CTS-21T. Water Environment Research Foundation.

Leersnyder, H. 1993. *The Performance of Wet Detention Basins for the Removal of Urban Stormwater Contaminants in the Auckland (NZ) Region.* M.S. Thesis. Dept of Environmental Science and Geography. University of Auckland.

Lenhart, James. 2007. *Evaluating BMP's Programs, Success and Issues.* Queensland Stormwater Industry Association, Annual State Conference. Sunshine Coast, Australia. Sept 2007 Keynote Address.

MacPherson, C. and B. Tonning. 2003. *Getting in Step: a guide to conducting watershed outreach campaigns.* Tetra Tech, Inc. EPA 841-B-03-002. U.S. EPA Office of Wetlands, Oceans and Watersheds. Washington, D.C.

Maestre, A., R. Pitt, and D. Williamson. 2004. "Nonparametric Statistical Tests Comparing First Flush and Composite Samples from the National Stormwater Quality Database." Stormwater and Urban WaterSystems Modelin. In: Models and Applications to Urban Water Systems. Vol. 12. Ed. W. James. CHI. Guelph, Ontario. Pp. 317-338.

Mahler, B. J., P.C. VanMetre and J. T. Wilson. 2004 (revised 2007). *Concentrations of polycyclic aromatic hydrocarbons (PAHs) and major and trace elements in simulated rainfall runoff from parking lots, Austin, Texas.* USGS Open File Report 2004-1208, U.S. Department of the Interior, U.S. Geological Survey, Reston, VA.

Malcolm, H.R., A.C. Beard, R.J. Burby, E.J. Kaiser, M.I. Luger, and R.G. Paterson. 1990. *Evaluation of the North Carolina Erosion and Sediment Control Program.* Raleigh, NC: Land Quality Section, Division of Land Resources, North Carolina Department of Environmental Health and Natural Resources.

National Oceanic and Atmospheric Administration (NOAA). 1988. *Dealing with Annex V – A Guide for Ports. U.S. Department of Commerce*, NOAA, National Marine Fisheries Service, Seattle, WA. NOAA Technical Memorandum NMFS F/NWR-23 as cited in *Guidance Specifying Management Measures for Sources of Nonpoint Pollution in Coastal Waters.* EPA 840-B-92-002.

Oberts, G. and R. Osgood. 1988. *Final Report on the Function of the Wetland Treatment Systems and the Impacts on Lake McCarrons.* Metropolitan Council. Publ. No. 590-88-095.

Ontario Ministry of the Environment. 2003. *Stormwater Management Planning and Design Manual 2003.* Appendix C. Simplified Design Approach. Ontario, Canada.

Oswald, G. and R. Mattison. 1995. "Monitoring the Effectiveness of Structural BMPs." Stormwater NPDES Related Monitoring Needs. ed. H. Torno. American Society of Civil Engineers. New York, NY.

Owens, L. K. 2002. *Introduction to Survey Research Design.* http://www.srl.uic.edu (accessed July 2007).

Pitt, R. and P. Bissonette. 1984. Bellevue Urban Runoff Program. Summary Report. Characterizing and Controlling Urban Runoff Through Street and Sewerage Cleaning. EPA/600/S2-85/038.

Pitt, R. and K. Parmer. 1995. *Quality Assurance Project Plan: Effects, Sources, and Treatability of Stormwater Toxicants*. U.S. EPA, Storm and Combined Sewer Program, Risk Reduction Engineering Laboratory. Cincinatti, OH.

Pitt, R. and J. Voorhees. 2002. *SLAMM, the Source Loading and Management Model.* Wet-Weather Flow in the Urban Watershed: Technology and Management, Edited by R. Field and D. Sullivan. CRC Press. Boca Raton, FL.

Pitt, R., R. Bannerman, S. Clark and D. Williamson. 2004a. "Sources of pollutants in urban areas (part 2) – Recent sheetflow monitoring." *Effective Modeling of Urban Water Systems, Monograph 13*. Edited by W. James, K. N. Irvine, E.A. McBean and R.E. Pitt.

Pitt, R. A. Maestre. 2004b. *The National Stormwater Quality Database (NSQD), version 1.1*. Department of Civil and Environmental Engineering, University of Alabama, Tuscaloosa, AL.

Pitt, R. and U. Khambhammettu. 2006. *Field Verification Tests of the UpFlow™ Filter. Small Business Innovative Research, Phase 2 (SBIR2) Report.* U.S. Environmental Protection Agency, Edison, NJ.

Pitt R., S.E. Clark, Y. Nara, R. Morquecho, and U Khambhammettu. 2008.

Sample processing to ensure subsample representativeness. *Journal of Environmental Engineering*, in press.

Polkwoska, Z., T. Górecki, and J. Namieśnik. 2002. "Quality of roof runoff waters from an urban region (Gdańsk, Poland)." *Chemosphere*, 49: 1275-1283.

Reeves, E. 1995. "Performance and Condition of Biofilters in the Pacific Northwest." Technical Note 30. *Watershed Protection Techniques* 1(3): 117-119.

Richter, B.D., J.V. Baumgartner, J. Powell, and D.P. Braun 1996. "A Method for Assessing Hydrologic Alteration Within Ecosystems." *Conservation Biology* 10:1163-1174.

Risse, M. and V. Jarrin. 2004. *Bacterial Source Tracking*. CSREES Southern Regional Water Quality Program. University of Georgia, Bulletin 1242-7.

Scanlin, J. and A. Feng. 1997. *Characterization of the Presence and Sources of Diazinon in the Castro Valley Creek Watershed.* Alameda Countywide Clean Water Program and Alameda County Flood Control and Water Conservation District, Oakland, CA.

Schueler, T. 2000. "Irreducible Pollutant Concentrations Discharged from Stormwater Practices." *The Practice of Watershed Protection.* Eds. T. Schueler and H. Holland. Center for Watershed Protection. Ellicott City, MD.

Schueler, T. R., C. Swann, T. Wright, and S. Sprinkle. 2004. *Pollution Source Control Practices – Version 1.0.* Manual 8 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection, Ellicott City, MD.

Schueler, T. and A. Kitchell. 2005. *Methods to Develop Restoration Plans for Small Urban Watersheds-Version 2.0.* Manual 2 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection. Ellicott City, MD.

Schueler, T., D. Hirschman, M. Novotney, and J. Zielinski. 2007. *Urban Stormwater Retrofit Practices-Version 1.0: A User's Manual.* Manual 3 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection. Ellicott City, MD.

Scheuren, F. 2004. *What is a Survey?* American Statistical Association.

Selbig, W. R. and R.T. Bannerman . 2007. *Evaluation of Street Sweeping as A Stormwater-Quality-Management Tool in Three Residential Basin in Madison, WI.* U.S. Geological Survey Scientific Investigations Report 2007-5156, 115pp. Available on-line (http://wi.water.usgs.gov/pubs/sci_invest.htm)

Selbig, W.R., R. Bannerman, and G. Bowman. 2007. "Improving the accuracy of sediment-associated constituent concentrations in whole water storm water samples by wet-sieving." *Journal of Environmental Quality*, 36: 226-232.

Sheehan. K. 2001. "E-mail survey response rates: A review." *Journal of Computer-Mediated Communication*, 6(2).

Shepp, D. 1995. *Petroleum Hydrocarbons in Urban Runoff: A Study Overview.* Prepared for Maryland Dept. of Environment. Department of Environmental Programs. Metropolitan Council of Governments. Washington, D.C.

Standard Methods for the Examination of Water and Wastewater. 19th edition. Water Environment Federation. Washington, D.C. 1995.

Smith, J., B. Hunt, and K. Bass. No Date. *Stormwater Best Management Practices (BMP) Monitoring Plan for the City of Charlotte and Mecklenburg County Stormwater Services.* Department of Biological and Agricultural Engineering, North Carolina State University.

Smith, E. P. 2002. *BACI design.* Encyclopedia of Environmetrics. Edited by El-Shaarawi, A.H. and W.W. Piegorsch. John Wiley and Sons, Ltd, Chichester.

Southern California Stormwater Monitoring Coalition (SMC). 2004. *Model Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California*; a Report from the Stormwater Monitoring Coalition's Model Monitoring Technical Committee.

Spetzman, J. no date. *Lawn Care and Water Quality: Finding the Balanc*e. Minnesota Department of Agriculture.

Steuer, J., W. Selbig, N. Hornewer and J. Prey. 1997. *Sources of Contamination in an Urban Basin in Marquette, MI and An Analysis of Concentrations, Loads and Data Quality.* US Geological Survey Water Resources Investigation Report 97-4242.

Strecker, E. 1995. "Constituents and Methods for Assessing BMPs." Stormwater NPDES Related Monitoring Needs. ed. H. Torno. American Society of Civil Engineers. New York, NY.

Strecker, E., M. Quigley, B. Urbonas, and J. Jones. 2004. "Stormwater Management; State-of-the-art in Comprehensive Approaches to Stormwater." *The Water Report; Water Rights, Water Quality & Solutions in the West.* Issue #6.

Strecker, E., W. Huber, J. Heaney, D. Bodine, J. Sansalone, M. Quigley, M. Leisenring, D. Pankani, and A. Thayumanauan. 2005. *Critical Assessment of Stormwater Treatment and Control Selection Issues.* Water Environment Research Foundation.

Steuer, J., W. Selbig, N. Hornewer, and J. Prey. 1997. *Sources of Contamination in an Urban Basin in Marquette, Michigan and an Analysis of Concentrations, Loads, and Data Quality*. U.S. Geological Survey, Water Resources Investigations Report 97-4242.

Taylor, A. and T. Wong. 2002. *Non-structural stormwater quality best management practices – a literature review of their value and life-cycle costs.* Technical Report 02/13. Cooperative Research Center for Catchment Hydrology. Victoria Environment Protection Agency.

Taylor, A., R. Curnow, T. Fletcher and J. Lewis. 2007. *Education campaigns to reduce stormwater pollutation in commercial areas: Do they work?* Journal of Environmental Management, 84: 323-335.

Taylor-Powell, E. and S. Steele. 1996. *Program Development and Evaluation, Collection Evaluation Data: Direct Observation*. G3658-5. University of Wisconsin-Extension. Program Development and Evaluation.

Taylor-Powell, E. 1998. Sampling. University of Wisconsin-Extension. Program Development and Evaluation.

Turner, M. 2005. *Leachate, soil and turf concentrations of nutrients from fertilizer; results from the Stillhouse neighborhood fertilizer leachate study*. Watershed Protection and Development Review Department, Environmental Resource Management Division, City of Austin, TX.

Underwood, A. J. 1992. "Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world." *Journal of Experimental Marine Biology and Ecology* 161: 145-178.

United States Department of Agriculture. 1996. *National Handbook of Water Quality Monitoring.* NRCS, Washington, DC. Available Online at: ftp://ftp.wcc.nrcs.usda.gov/downloads/wqam/wqm1.pdf

University of Alabama (UA) and Center for Watershed Protection (CWP). 2008. *National Stormwater Quality Database, Version 3*. University of Alabama. Online: http://unix.eng.ua.edu/~rpitt/Research/ms4/mainms4.shtml

Urbonas, B. 1995. *Parameters to Report with BMP Monitoring Data.* Stormwater NPDES Related Monitoring Needs. Edited by H. Torno. American Society of Civil Engineers. New York, NY.

U.S. Environmental Protection Agency (EPA). 1983. *Results of the Nationwide Urban Runoff Program, Final Report.* PB84-185552, December 1983.

U.S. Environmental Protection Agency (EPA). 1982. *Handbook for Sampling and Sample Preservation of Water and Wastewater.* Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH, EPA 600/4-82/029.

U.S. EPA. 2003. *Getting In Step: A Guide for Conducting Watershed Outreach Campaigns*. EPA 841-B-03-002.

U.S. EPA. 2005. *Stormwater Phase II Final Rule.* Public Education and Outreach Minimum Control Measure. Office of Water. EPA 833-F00-005.

U.S. Geological Survey (USGS). 2006. *National Field Manual for the Collection of Water-Quality Data.Chapters A4. Collection of Water Samples.* Techniques of Water-Resources Investigations, Book 9, Handbooks for Water-Resources Investigations. http://water.usgs.gov/owq/FieldManual/chapter4/pdf/Chap4_v2.pdf

Waschbusch, R.J., Selbig, W.R. and Bannerman, R.T. 1999. *Sources of Phosphorus in Stormwater and Street Dirt from Two Urban Residential Basins in Madison, Wisconsin, 1994-1995*. U.S. Geological Survey, Water Resources Investigations Report 99-4021

Waschbush, R.J. 2003. *Data and Methods of a 1999-2000 Street Sweeping Study on an Urban Freeway in Milwaukee County, Wisconsin.* Open File Report 03-93. U.S. Department of the Interior, U.S. Geological Survey.

Watershed Environment Research Federation. 2000. *Tools to Measure Source Control Program Effectiveness.* Alexandria, VA.

Watt, J. H. 1997. "Using the internet for quantitative survey research." *Marketing Research Review*, July.

Winer, R. 2000. *National Pollutant Removal Database for Stormwater Treatment Practices*. Second edition. Center for Watershed Protection. Ellicott City, MD.

Wright, T., C. Swann, K. Cappiella, T. Schueler. 2004. *Unified Subwatershed and Site Reconnaissance: A User's Manual.* Manual 11 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection. Ellicott City, MD.

Appendix

A

Selecting Samples and Determining Sample Size for Stormwater Monitoring

A description of various methods to select samples and determine sample size is provided in this Appendix. Sample size estimation methods provided in this Appendix are applicable to Study Designs 1, 2, 3 and 6 and are for study designs that are collecting individual and paired water quality samples.

The statistical basis for this approach is required to justify the allocation of limited resources and to determine the sampling effort needed to accomplish project objectives.

In many cases, certain elements of a monitoring program require much more time and money than other elements of the program. The approach and tools given in this Appendix enable project managers to balance project resources and scope with expected outcomes. It can be devastating to project conclusions if the needed sample size was not properly estimated and therefore obtained during the course of the monitoring study. These tools enable one to better plan and conduct a sampling program to minimize this possibility.

Selection of Sampling Plan

A sampling plan identifies the method by which the samples are to be selected. A sample is a proportion of the population where information (usually average values, totals, ranges, etc.) about a large population may be inferred. Sampling plans have been well described in the environmental literature. Box A1 describes four main categories, plus subcategories, of sampling plans. Probability based sampling is a commonly used sampling method for environmental studies. A sampling plan must be selected before proceeding to the next step of sampling - determining the sample size.

Determining Sample Size

Box A1. Sampling Plans (adapted from Burton and Pitt, 2002)

Haphazard Sampling – Samples are taken in a haphazard manner, usually at the convenience of the sampler when time permits.

Judgment Sampling – When only a specific subset of the total population is to be evaluated, with no desire to obtain "universal" characteristics.

Probability Sampling

Simple Random Sampling – Samples are taken randomly from the complete population.

Stratified Random Sampling – Samples are randomly obtained from several population groups that are assumed to be internally more homogeneous than the population as a whole.

Multistage Sampling – An initial sampling effort is used to examine major categories of the population that may be divided into separate clusters during later sampling activities.

Cluster Sampling – Specifically targeting specific population units that cluster together.

Systematic Sampling – Evenly spaced samples are collected for an extended time.

Search Sampling – Used to find specific conditions where prior knowledge is available and stresses areas thought to have a greater probability of success.

An important aspect of any monitoring program is assurance that the samples collected represent the conditions to be tested and that the number of samples to be collected are sufficient to provide statistically relevant and informative conclusions. Unfortunately, sample numbers are most often not based on a statistically-based process and follow traditional "best professional judgment," or are resource driven. If resources (e.g. funding and/or personnel) limit the number of samples being collected (which they invariably do), a power analysis is recommended to determine what may be accomplished with the available resources. A power analysis uses the methods described below to determine the number of samples that can be collected and the significance that may be achieved based on available resources.

To estimate sample size, some knowledge about the water quality parameters to be monitored is needed such as the expected mean and standard deviation. This information can come from the study sites themselves from previous monitoring studies or allocating resources to obtain samples prior to beginning the current monitoring program. The mean and standard deviation of water quality parameters may also be estimated using existing databases such as the National Stormwater Quality Database (NSQD). These two statistical parameters, mean and standard deviation, are needed to calculate the coefficient of variation (COV) to approximate sample size estimated using the graphics provided in this Appendix (see Figures A1 and A2). Spreadsheets such as Excel and other statistical programs may be used to calculate means and standard deviations among many other statistical parameters.

Statistical estimates of sample size are recommended for stormwater monitoring and are relatively straight-forward to estimate based on the graphical plots provided in Figures A1 and A2. Additional plots are provided in Burton and Pitt (2002). A separate estimate of sample size for each monitored parameter is suggested given the differences amongst water quality parameters. That is, the variability in total suspended solids (TSS), total nitrogen, nitrate-nitrogen, or lead will likely differ and may lead to a different estimate in the number of samples to be collected. For multiple monitoring sites, an equal number of samples should be taken at all sampling locations if comparing station data (EPA 1983). If the monitoring program is based on paired sampling, the samples at the two comparison sites should be collected at the "same" time, for example, allowing for much more powerful paired statistical comparison tests.

The plots provided in this Appendix are based on equations presented in Burton and Pitt (2002). The equations are provided here for information purposes only.

Sample size estimate for individual samples to characterize conditions

For most stormwater monitoring circumstances Figure A1 can be used to graphically estimate the number of samples. The equation is provided below for those who have a more advanced understanding of statistical methods. The sample size estimates provided in Figure A1 are based on the following equation that uses the statistical parameters such as the allowable error, the variance of the observations, and the degree of confidence and power. The basic equation is based on Cameron (undated) as cited in Burton and Pitt (2002) as follows:

$n = [COV(Z_{1-\alpha} + Z_{1-\beta})/(error)]^2$ [Equation 1]

where:

n = number of samples needed

α= false positive rate (1-α is the degree of confidence. A value of α of 0.05 is usually considered statistically significant, corresponding to a 1-α degree of confidence of 0.95, or 95%.) (See Box A2)

β= false negative rate (1-β is the power. If used, a value of β of 0.2 is common, but it is frequently ignored, corresponding to a β of 0.5.) (See Box A2)

$Z_{1-\alpha}$ = Z score (associated with the area under the normal curve) corresponding to 1-α. If α is 0.05 (95% degree of confidence), then the corresponding $Z_{1-\alpha}$ score is 1.645 (from standard statistical tables).

$Z_{1-\beta}$= Z score corresponding to 1-β value. If β is 0.2 (power of 80%), then the corresponding $Z_{1-\beta}$ score is 0.85 (from standard statistical tables). However, if power is ignored and β is 0.5, then the corresponding $Z_{1-\beta}$ score is 0.

error = allowable error, as a fraction of the true value of the mean

COV = coefficient of variation (sometimes noted as CV), the standard deviation divided by the mean (data set assumed to be normally distributed.)

The equation is for data that has a normal distribution (the classic bell-curve shape), which is rare for most stormwater data. Therefore the sample size estimates are only approximate and used as a guide to scope the monitoring project. The coefficient of variation (COV) is a measure of variability in the data. If the coefficient of variation (COV) values are low (less than about 0.4), then there is likely no significant difference

in the predicted sampling effort. A higher COV in the data is desired in order to more readily detect significant differences in the sampling effort. In most stormwater monitoring situations it is likely to encounter COV values closer to 1.

Figure A1 is a plot of Equation 1 to approximate the number of individual samples needed for a 95% degree of confidence ($\propto = 0.05$) and a power of 80% ($\beta = 0.2$). For example, if an allowable error of 20% (0.2) is deemed acceptable and the COV of the monitoring parameter is 0.7, then approximately 45 samples are needed, compared to 31 samples if the variability in the water quality parameter is lower (COV = 0.5). Sample size increases as the COV increases and error decreases. The appropriate selection of these statistical parameters to determine sample size is needed as it affects the timeframe for the monitoring study (and we all know that time is money).

Box A2. The "alpha" and "beta" of statistical errors

Errors in decision making are usually divided into type 1 (α: alpha) and type 2 (β: beta) errors:

α (alpha) also known as a Type 1 error is a false positive. A false positive is assuming something is true when it is actually false. An example would be concluding that a tested water was adversely contaminated, when it actually was clean. The most common value of α is 0.05 (accepting a 5% risk of having a type 1 error), although other values may be appropriate for specific project objectives and stages. Confidence is 1-α, or the confidence of not having a false positive.

β (beta) also known as a Type 2 error is a false negative, or assuming something is false when it is actually true. An example would be concluding that a tested water was clean when it actually was contaminated. If this was an effluent, it would therefore be an illegal discharge with the possible imposition of severe penalties from the regulatory agency. In most statistical tests, β is usually not directly considered (if ignored, β is 0.5), but is assumed to be considered during the experimental design phase with adequate samples collected to control the false negative rate. A typical value of β is 0.2, implying accepting a 20% risk of having a Type 2 error. Power is 1-β, or the certainty of not having a false negative. Again, other levels of power may be appropriate for the specific project objectives.

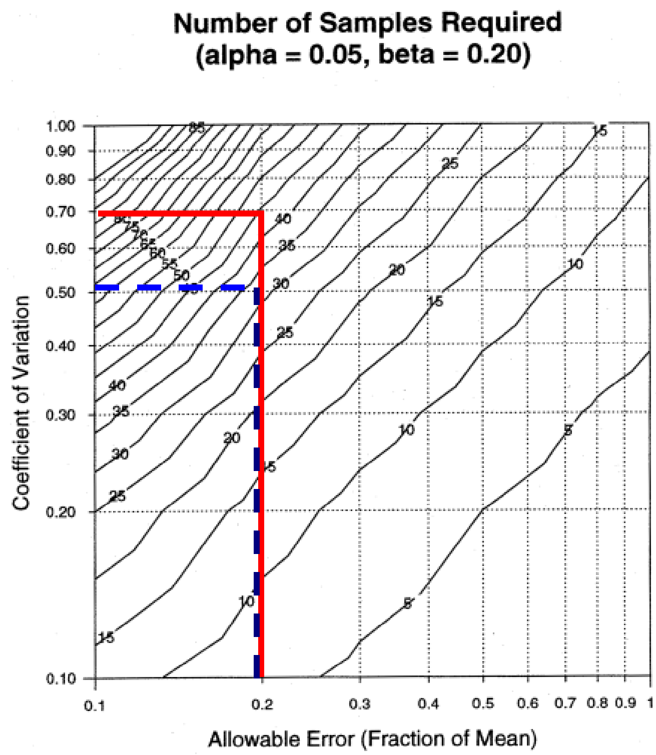


Figure A1. Sample size estimates for individual samples for confidence of 95% (α=0.05) and a Power of 80% (β=0.20) (Pitt and Parmer 1995)

Sample size estimate for paired samples

The comparison of paired data sets is commonly used when evaluating the differences between two situations (locations, times, practices, etc.). This method to estimate sample size for paired samples is applicable to Study Designs 1, 3 and 6. The International BMP database (http://www.bmpdatabase.org) may be used to get information on influent and effluent concentrations that can be used for Study Design 3 sample size estimates, while estimates of concentration from different land use catchments may be acquired through the NSQD to estimate the sample size needed for Study Design 5.

A related equation to Equation 1 can be used to estimate the needed samples for a paired comparison (Cameron, undated) as cited in Burton and Pitt (2002) as follows:

$$n = 2\,[(Z_{1-\alpha} + Z_{1-\beta})/(\mu_1 - \mu_2)]^2\sigma^2 \qquad \text{[Equation 2]}$$

where:

α = false positive rate (1-α is the degree of confidence. A value of α of 0.05 is usually considered statistically significant, corresponding to a 1-α degree of confidence of 0.95, or 95%)

β = false negative rate (1-β is the power. If used, a value of β of 0.2 is common, but it is frequently ignored, corresponding to a β of 0.5.)

$Z_{1-\alpha}$ = Z score (associated with area under normal curve) corresponding to 1-α

$Z_{1-\beta}$ = Z score corresponding to 1-β value

μ_1 = mean of data set one

μ_2 = mean of data set two

σ = standard deviation (same for both data sets, same units as μ. Both data sets are also assumed to be normally distributed.)

Figure A2 from Burton and Pitt (2002) plots the results of Equation 2 using the coefficient of variation (COV) and difference in sample set means. The approximate number of sample pairs needed is based on a degree of confidence of 95% (α of 0.05), and a power of 80% (β of 0.2). Similar to the other equation, the sample size is only approximate, as it requires that the two data sets be normally distributed and have the same standard deviations. The COV is the average for the two datasets (e.g. average COV for influent and effluent quality of a STP). Again, if the COV values are low (less than about 0.4), then there is probably no real difference in the predicted sampling effort.

A larger number of paired samples are needed the higher the COV, or the smaller the difference between sample set means. For example, a higher COV, or greater variability in stormwater concentration, would require more sample pairs in order to characterize or capture the range in stormwater quality. Figure A2 also illustrates that with a greater difference in influent and effluent concentrations, fewer samples need to be compared to detect a difference. Using Figure A2 to illustrate, if the COV is equal to 1 and the difference in sample set means is 50%, approximately 75 paired samples are needed, compared to only 20 sample pairs if the difference in sample set means is 80%. Burton and Pitt (2002) contains similar plots for many other combinations of power, confidence and expected differences that may be used in a power analysis to determine the sample size needed to satisfy the monitoring objectives.

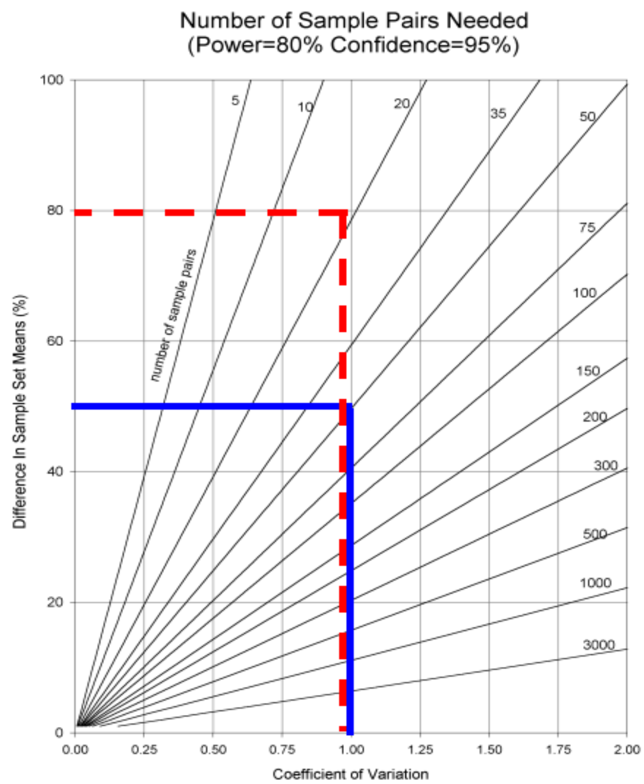


Figure A2. Sample Effort Needed for Paired Testing (Power of 80% and Confidence of 95%) (Pitt and Parmer, 1995)

Resources

ASCE and U.S. EPA. 1999. Development of Performance Measures. Task 3.1 – Technical Memorandum. Determining Urban Stormwater Best Management Practice (BMP) Removal Efficiencies. Available at, http://www.bmpdatabase.org/Docs/task3_1.pdf

Burton, A., and R. Pitt. 2002. Stormwater Effects Handbook: a Toolbox for Watershed Managers, Scientists, and Engineers. Lewis Publishers. New York, NY. Available at http://unix.eng.ua.edu/~rpitt/Publications/Publications.shtml

EPA. 1983. *Methods for Chemical Analysis of Water and Wastes*, EPA-600/4-79-020, U.S. Environmental Protection Agency, Cincinnati, Ohio..

International Stormwater (NSW) Best Management Practices (BMP) Database. 2006. Available online at: http://www.bmpdatabase.org

U.S. EPA's Urban BMP performance evaluation tool Available online at: http://cfpub.epa.gov/npdes/stormwater/urbanbmp/bmpeffectiveness.cfm

National Pollutant Removal Performance Database available online at:
http://www.cwp.org/PublicationStore/special.htm

R. Pitt and K. Parmer. 1995. *Quality Assurance Project Plan (QAPP) for EPA Sponsored Study on Control of Stormwater Toxicants*. Department of Civil and Environmental Engineering, University of Alabama at Birmingham.

Appendix

B

Example Statistical Analyses to Evaluate Monitoring Data

Central to any monitoring program is the organization and analysis of the data generated. Fortunately, software programs and packages (e.g. Excel, SAS, Sysstat, Sigma Stat) are available to help organize datasets and provide relatively painless procedures to analyze the monitoring data. So, no need to worry if you haven't opened your statistical textbook in the past ten years, or more. This Appendix material provides information on the basic steps to analyze data to determine the 'meaning' of the data generated using the study designs described in this manual. For example, statistical tests can determine the performance of a STP by comparing influent and effluent quality of the STP (Study Design 3), or whether differences in water quality may be attributed to treatment in a paired watershed study (Study Design 5). Readers are encouraged to review materials in the Resources section of this Appendix to provide more detailed description of the procedures presented in this Appendix. For example Grabow *et al.* (1998, 1999) provide excellent step-by-step instructions on the statistical analyses using Excel and SAS, while Burton and Pitt (2002) provide more detailed description of the methods described here along with additional statistical methods and their applications.

Getting Started

The first step in statistical analyses is to familiarize yourself with the data. This is achieved by having the data well organized to support the analyses. The organization of the data is related to the monitoring study objective. For example, if the objective of the monitoring study is to determine the effect of an STP on effluent quality, the rows of data should be paired observations of influent and effluent concentrations. Additional information that is useful to data management and analyses are: the date of data collection, flow, and precipitation as separate column entries. A column to provide comments to annotate individual observations can provide additional information to aid analyses.

Exploratory Data Analyses

Exploratory data analysis provides information to determine if the data has a normal distribution and if parametric statistical procedures can be used. Alternatively, the data may not have a normal distribution and non-parametric procedures may be needed, or the data needs to be transformed to approximate a normal distribution. In most cases, stormwater data is rarely normally distributed and requires the data to be transformed (e.g., transformed to a base 10 logarithm) such that it approximates a normal distribution. For example, using the real data values, the "log10" or natural log "ln" function in Excel can be used to transform the values. Data transformation is preferred so that parametric statistical procedures can be used, as they are more robust (provide more information) compared to non-parametric statistics.

The 'normality' of the data is commonly determined based on a measure of skewness as an index for the distribution of the data. A normal distribution is not 'skewed' and is illustrated by a bell-shaped curve that is symmetric about the mean (similar distribution of the data above and below the mean value of the dataset). A dataset is considered skewed if the data is not evenly distributed about the mean. Typically, stormwater data is positively skewed where there are many relatively low values and very few high values (see Figure B1.) (Grabow *et al.* 1998, 1999). Data transformations are therefore needed if parametric statistical analyses are planned.

Non-parametric statistics can be a useful starting point for analyses to first determine if there is a significance difference between two datasets (e.g. control vs. treatment in a paired watershed study, influent and effluent in a BMP).

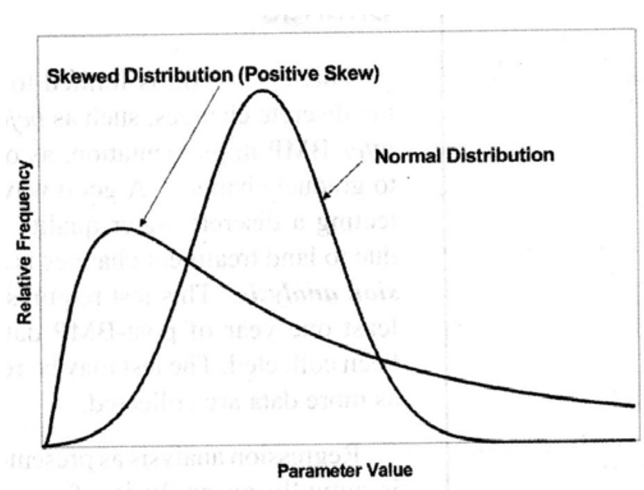


Figure B1. Illustration of a normal and skewed distribution (from Grabow et al. 1998)

Graphical analyses

Graphical analyses using simple plots are an effective way to visually assess data. Figure B2 is a scatterplot of observed influent concentrations vs. the effluent concentrations for suspended solids under actual rain conditions. This plot shows generally large reductions in TSS concentrations for most events. For example, the data points in Figure B2 are mostly below the solid line that shows the effluent concentrations are lower than the influent concentrations. If the data plotted along the "1-to-1" line this would indicate no difference between the influent and effluent TSS concentrations; that is the influent equals the effluent.

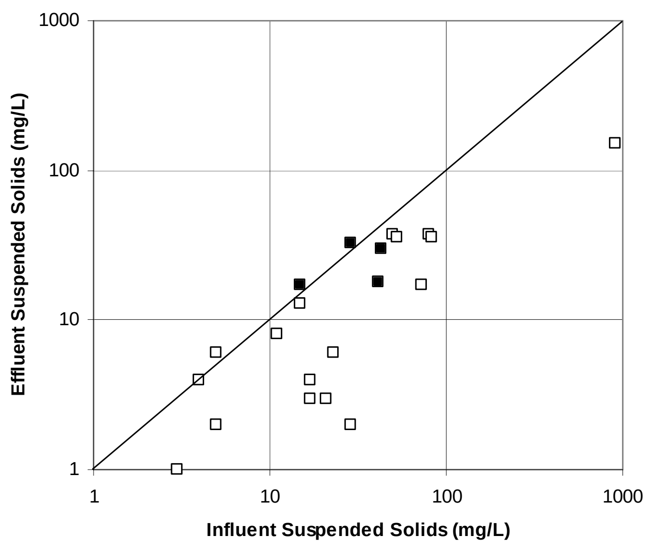


Figure B2. Scatterplot of observed influent and effluent suspended solids concentrations (filled symbols are events that had minor filter bypasses)

Box-and-whisker plots

Box and whisker plots are a useful tool to visually compare the distribution from datasets and to quickly determine if significant differences exist. Concentration, loads and flow data are useful data to compare using this graphical statistical technique. The box and whisker plots summarize the following features from the data:

Median Efficiency = 50% of the values are above or below this value

1st quartile or 25th Percentile = 25% of the values fall at or below this value

3rd quartile or 75th Percentile = 75% of the values fall at or below this value

Upper inner fence = Highest value

Lower inner fence = Lowest value

Upper and Lower 95% confidence level = Value that fall within the 95% confidence value of an alpha (α) value of 0.05.

The log-transformed data is used in the analyses, or other data transformations as needed to approximate a normal distribution. A box-and-whisker plot for each group of data (e.g. influent, effluent, outfall of control watershed and outfall of treatment watershed) is prepared using a statistical software program. A comparison of multiple, or grouped box plots can be made to visually assess significant differences. Figure B3 is an example of a box and whisker plot. If the box and whisker plot were illustrating the pollutant concentration data for the influent and effluent of an STP a significant difference would be determined by looking at the degree of overlap between the respective confidence levels of the median. A significant difference exists if there is no overlap between the confidence levels of the two box and whisker plots.

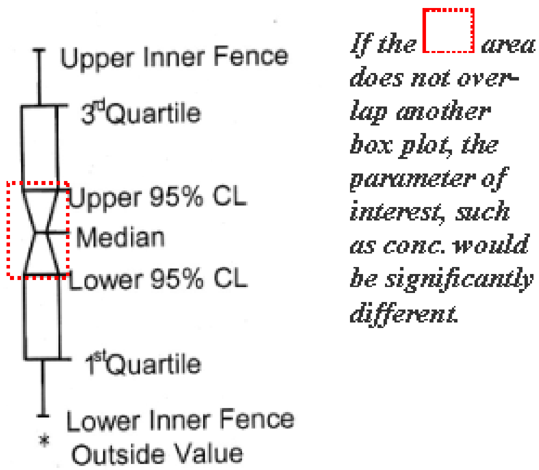


Figure B3. Example box-and-whisker plot

Test Statistics

To determine the significance of the data, parametric or nonparametric statistical techniques may be used. Table B1 provides a summary of commonly used statistical tests. Parametric test statistics have more rigorous assumptions to be satisfied for the test to be "trustworthy" compared to more relaxed conditions for nonparametric tests. Nonparametric tests use rank-order of the individual data/observations rather than the absolute value of the data itself. That is, the values are ranked in ascending order. Statistical software packages and programs are available to perform these analyses and are described more fully in the Resources section listed in the end of Appendix B.

Table B1. Commonly used statistical tests		
Purpose of Test	**Parametric**	**Nonparametric**
Compares 2 independent samples (e.g. influent and effluent)	Student T- test	Mann-Whitney U test; Wilcoxon rank-sum test
Examines sets of differences	Paired T-test	Wilcoxon signed-rank test, Sign Test
Example use of Signed ranked test: Pitt and Khambhammettu (2006) used a nonparametric sign test (such as the Wilcoxon Signed Rank Test) to determine if the influent concentration was significantly different compared to the effluent concentration for an STP. For the TSS data, the probability values was less than 0.01 (P < 0.01), indicating with greater than 99% confidence that the influent does not equal the effluent concentrations. As an aside, if the p-value was 0.03 (greater than 0.01 but less than 0.05), this would indicate the influent concentration does not equal the effluent concentration at a 95% confidence level (α=0.05).		

Regression Analyses

Exploratory variables, also referred to as covariates, are variables collected during the monitoring program that explain the variability in the data. Regression analysis is used to determine what variables best explain the variability, or help understand a cause and effect relationship. Essentially, you are trying to quantify how one variable (independent variable) affects, or covaries with another (dependent variable). Examples of independent and dependent variables are listed in Table B2 for various study designs. Regression analysis uses the relationship between the monitoring (observed) data of the independent and dependent variables in order to predict other or future values of the dependent variable. A linear regression equation is expressed as follows,

$$Y = \beta x + b,$$

Where Y is the predicted value based on observed values of "x" the independent variable, β is the slope of the line and 'b' is the y-intercept (where the regression line crosses the y-axis).

Regression analysis may be used to predict effluent concentration for a STP based on a defined relationship that is established from the monitoring data (e.g., influent and effluent concentration as in Study Design 3). In a paired watershed study the regression analyses would help determine if the treatments applied to a watershed (e.g. use of STPs) had an impact on water quality compared to the control watershed where no best management practices, but similar land use was present (Study Design6). Analysis of Variance (ANOVA) or Analyses of Covariance (ANCOVA) is a commonly used technique to define the regression equation and its significance. The basic idea in an ANOVA is that it not only provides coefficients (e.g. β) to define the regression equation and its significance, but also to assess the variation in the dataset. Information from an ANOVA or ANCOVA can determine if the total variation of a dataset can be attributed to a specific source or cause of variation, or if it is just attributed to chance and there is no significant relationship between the explanatory variables, or covariates.

Table B2. Example variables in three study designs		
Study design	**Independent variable (X)**	**Dependent variable (Y)**
Paired watershed or site *Applies to Study Designs 3, 5, or 6*	Water quality parameter of interest collected from control watershed	Same water quality parameter of interest but collected from the treatment watershed
Longitudinal (Upstream/downstream, or inflow/outflow) *Applies to Study Design 3*	Water quality parameter of interest collected upstream or inflow	Same water quality parameter but collected downstream or outflow
Source: Grabow *et al.* 1998		

An Example using ANOVA to determine the performance of a STP

Using data from Pitt and Khambhammettu (2006), regression equations were fitted to determine the significant differences between influent and effluent from an STP using ANOVA. In all cases, the data needed to be log-transformed in order to obtain proper residual behavior, or constant variance. That is, there is no trend or pattern in the residual, where the residual is the difference between the observed value and the predicted 'Y' value. The 'proper' residual behavior is to have the residual evenly scattered about the 'zero' line. Figure B4 illustrates the output from an ANOVA analyses. Using data from Pitt and Khambhammettu (2006) for TSS, the following fitted regression equation was found to be very significant, according to the ANOVA analyses on a $\log_{10}$ scale:

Effluent Suspended Solids, log mg/L = 0.730 * (Influent Suspended Solids, log mg/L), [Equation 1]

Figure B4. Regression Statistics on Observed Influent vs. Effluent Suspended Solids, log mg/L using ANOVA

Multiple R	0.94
R Square	0.89
Adjusted R Square	0.85
Standard Error	0.37
Observations	24

ANOVA

	Degrees of freedom	Sum of squares	Mean squares	F-statistic	Significance F-statistic
Regression	1	25.4	25.4	187	3.11E-12
Residual	23	3.12	0.136		
Total	24	28.55			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
X Variable 1	0.730	0.053	13.7	1.56E-12	0.620	0.841

There is a lot of information associated with an ANOVA or ANCOVA table including a few basic elements to determine the significance of the relationship: the Sum of squares, F statistics, p-value, and coefficient. Table B3 provides a summary of the output and its interpretation. The output in Table B3 is based on a single dependent variable (X). Where more than one dependent variable or factors are considered an ANCOVA may be used. Grabow et al. (1998, 1999) provide easy to follow instructions using ANCOVA for a paired watershed study. Figure B4 illustrates the plotted regression equation in pink diamonds (Equation 1) along with the observed or monitoring effluent and influent concentration (blue diamonds). Figure B5 shows four different ways to visualize residuals of a regression where the residual is the difference between the observed and fitted regression equation. A proper pattern of residual would show no bias or trend in this plot. If a pattern does emerge, it would indicate a variable may not have been accounted for that can significantly affect the statistical relationship.

Table B3. Summary output from an ANOVA table

Output	Value	Interpretation
Sum of Squares Regression Total	25.4 28.55	Provides information on how much of the total variation can be attributed to the regression. Calculating, 25.4/28.55 = 0.889. This value is equivalent to the r-square value. Roughly 88.9% of the total variation effluent concentration is explained by the regression equation using the influent concentration.
F-statistic	187	The F-statistic is the ratio of the mean squares. A large F-statistic implies that the amount of variation explained by the regression is large in comparison with the error (which is not explained by the regression). For this example, the F-statistic is 187 (25.4/0.136) with a very low significance value of 3.11E-12. Since 3.11E-12 < 0.05 it may be concluded that the regression is significant at the 95% confidence level with an associated α of 0.05, or 0.01 for a 99% confidence level.
X variable 1 (e.g. influent concentration)	0.73	This is the coefficient of the independent variable, the slope term (β).
p-value	1.56E-12	The p-value is the probability value to assess the significance of the coefficient. As the p-value is much less than an α of 0.05, or 0.01 it may be concluded that the slope term of the equation is also highly significant

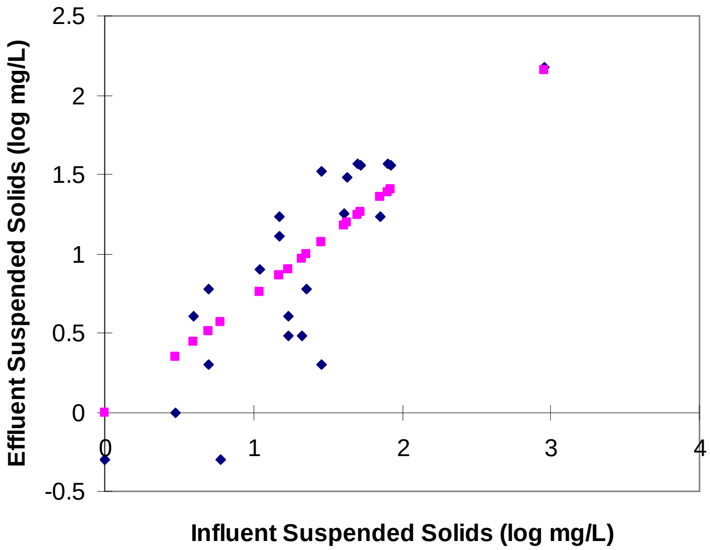


Figure B5. Fitted regression equation and data points for influent and effluent suspended solids

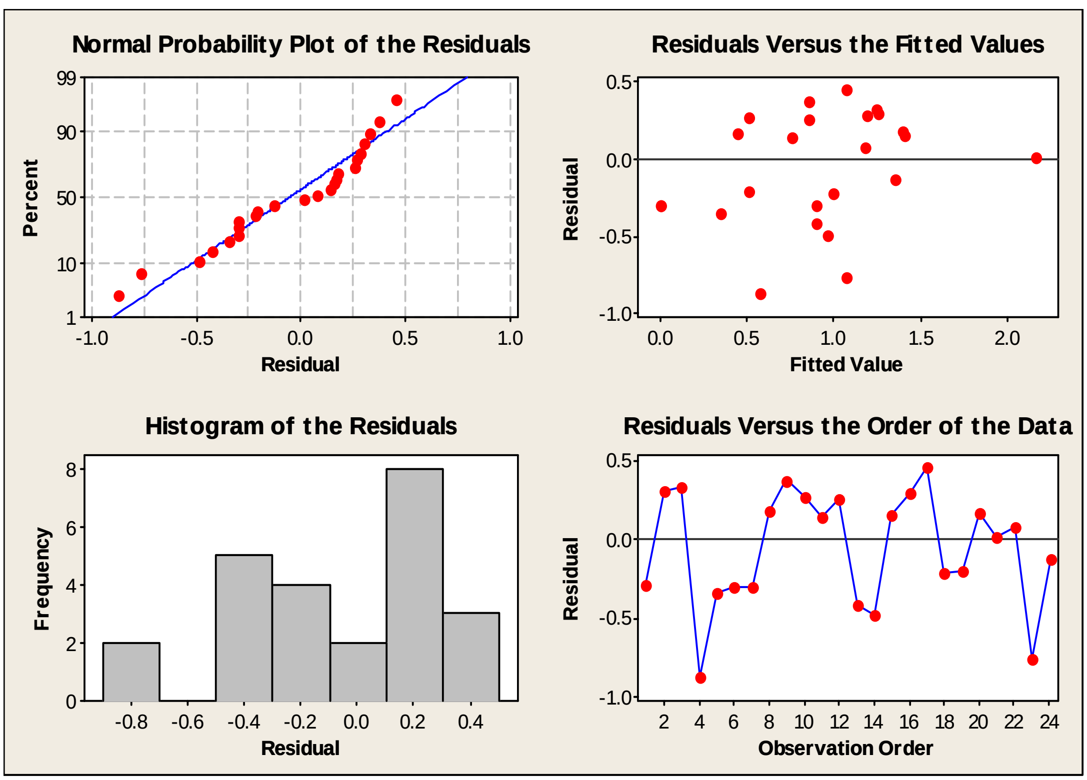


Figure B5. Residual analyses of fitted equation for suspended solids influent vs. effluent

Using ANCOVA for a Paired Watershed Study

The output for an ANCOVA is similar to an ANOVA table illustrated in Table B3 with the exception that there will be additional "x-variables" listed and their associated statistics. Grabow et al. (1998, 1999) provide instructions on how to organize the data and perform ANCOVA for a paired watershed study. The resources section provides a link to download these reports.

Using the Chi-Square Test

The Chi Square test is used to test the independence of two categorical variables, where two datasets are compared to determine if their distributions are statistically different or not. The Chi-Square test only tests for the significance of this difference between two variables, and not the magnitude of this difference (more or less). Statistical software or spreadsheet programs have the ability to automatically compute a Chi-Square test statistic. Although, an Excel spreadsheet requires additional user input (but more on this later). The following is an overview of how data is organized to compute the Chi-Square test statistic and its interpretation. Review of material provided in the resource section can provide a more detailed description of the Chi-Square test.

Data organization

The results of a homeowner lawncare survey are presented in Table B4. The survey was part of a paired watershed study that was designed to monitor the impact of an educational campaign to reduce the use of fertilizers in residential catchments. A survey was administered during a calibration period where baseline lawn management practices by homeowners were documented. The survey was repeated following an education campaign in the same catchment. Results indicate that 38% of residents fertilize their lawns, a decrease from 12% from the calibration period. The survey also found that 36% do not fertilize their lawn, an increase from the calibration period. The Chi-square test statistic is used to determine if this difference is significant. To begin, the data from the survey is organized into a contingency table as illustrated in Tables B5a-b.

Table B4. Percent of homeowners who fertilize and do not fertilize lawns based on survey results

	Calibration period	Treatment period
Fertilize	50	38
Do not fertilize	20	36

Table B5a. Example of how to calculate expected values using a 2x2 contingency table			
Variable 1	Data type 1	Data type 2	Total
Fertilize	**A**	**b**	**a** + **b**
Do not fertilize	**C**	**d**	**c** + d
Total	**a+c**	**B+d**	**a + b + c + d = N**
Table B5b. Results of survey entered into a 2x2 contingency table.			
Treatment catchment	Pretreatment	Treatment	Total
Fertilize	**50**	**38**	88
Do not fertilize	**20**	**36**	56
Total	70	74	144

Calculating the Chi-Square Test Statistic (χ^2)

There are 5 basic steps to calculate the χ^2 statistic.

1. State the hypotheses and level of confidence to determine significance of results.

2. Calculate the expected values using the values in the contingency table

3. Determine the degrees of freedom.

4. Calculate the χ^2

5. Compare the χ^2 and compare to table of values at predetermined level of confidence.

The Chi-square test statistic is used to determine if the two variables are independent. That is, what is the likelihood that fertilizing a lawn is the same for the calibration period as the treatment period. This is the *null hypothesis*. The alternative hypothesis to be tested is that the likelihood of fertilizing a lawn is not the same for the pretreatment and treatment period. The objective is to reject the null hypothesis and accept the alternative hypothesis.

Calculate the 4 expected values and enter these values into a table (Table B6).

(Total number of residents who fertilize during calibration period x Total number of residents in calibration period)/(total "n" for the table).

(Total number of residents who do not fertilize during the calibration period x Total number of residents in calibration period)/(total "n" for the table).

(Total number of residents who fertilize during the treatment period x Total number of residents in treatment period)/(total "n" for the table).

(Total number of residents who do not fertilize during the treatment period x Total number of residents in treatment period)/(total "n" for the table).

Table B6. Expected values entered into a 2x2 contingency table		
Treatment catchment	Pretreatment	Treatment
Fertilize	**42.8**	**45.2**
Do not fertilize	**27.2**	**28.8**

Determine the degrees of freedom by,

(number of columns -1) multiplied by (number of row -1).

In the example, the contingency table is two columns by two rows, or 2x2. The degrees of freedom is (2-1) x (2-1) that is equal to 1.

The χ^2 is calculated by,

sum of the ((expected value – observed value)^2/expected value).

For example,

$\chi^2 = (50\text{-}42.8)^2/42.778 + (38\text{-}45.2)^2/45.222 + (20\text{-}27.2)^2/27.222 + (36\text{-}28.8)^2/28.8$

$\chi^2 = 6.101$

The value of 6.101 is used to compare values in a standard statistical table of critical values for Chi-square test statistics with 1 degree of freedom. In this example, 6.101 exceeds the critical value in the table for probability value or "alpha (α)" of 0.05 of 3.841. In Excel, the "chi-test" statistical function provides a probability value that is compared to the defined probability value. In this example, the p-value for the dataset provided by Excel is 0.013 which is less than 0.05. Therefore, it may be concluded that there are differences in fertilizing between the pretreatment and treatment periods at the 95% level of confidence and the null hypothesis can be rejected.

Resources

Burton, A., and R. Pitt. 2002. Stormwater Effects Handbook: a Toolbox for Watershed Managers, Scientists, and Engineers. Lewis Publishers. New York, NY. Available at http://unix.eng.ua.edu/~rpitt/Publications/Publications.shtml

Clausen, J. C., and J. Spooner. 1993. *Paired Watershed Study Design*. 841-F-93-009. U.S. EPA Office of Water. Washington, DC.

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1998. Detecting Water Quality Changes Before and After BMP implementation: Use of a Spreadsheet for Statistical Analysis. In: NWQEP Notes, Number 92, November, 1998. Available at http://www.bae.ncsu.edu/programs/extension/wqg/issues/pubindex.html (accessed February 2008)

Grabow, G.L., J. Spooner, L.A. Lombardo, and D.E. Line. 1999. Detecting Water Quality Changes Before and After BMP implementation: Use of SAS for Statistical Analysis. In: NWQEP Notes, Number 93, January, 1999. Available at http://www.bae.ncsu.edu/programs/extension/wqg/issues/pubindex.html (Accessed February 2008)

Pitt, R. and U. Khambhammettu. *Field Verification Tests of the UpFlow™ Filter. Small Business Innovative Research, Phase 2 (SBIR2) Report.* U.S. Environmental Protection Agency, Edison, NJ. 275 pages. March 2006.

Statistics explained online:

http://www.statsoft.com/textbook/stathome.html

http://davidmlane.com/hyperstat/

Appendix

C

Costs Associated with Monitoring Studies: Analysis of Water Quality Parameters and Unit Costs for Education and Outreach Programs

A summary of estimated sample costs for a set of water quality parameters is provided in Table C1. The costs are for chemical analysis only and do not include sample supplies, shipping or other materials used to preserve the sample, or staff time to collect the samples. The information was compiled based on a review of publications, monitoring studies and consultations with stormwater practitioners. These costs largely reflect a contract, or outside lab doing the analyses. Costs will vary based on the analytical method used, which should be specified in the QA/QC plan for the study design. For example, the BOD-5 method for BOD analysis is much less expensive compared to titration methods. Costs may be discounted if a bundle or group of parameters were to be analyzed. If the chemical analyses can be done in-house using existing facilities and certified, trained lab technicians, the costs would be lower, or would reflect staff time to do the analyses.

Table C1. Example Cost for 14 Water Quality Parameters

Parameter	Per Sample Costs
TSS	$10-15
BOD	$30-35
COD	$25-30
CU	$15-25
Pb	$15-25
Zn	$10-25

Table C1. Example Cost for 14 Water Quality Parameters	
Parameter	**Per Sample Costs**
TP	$22-35
TKN	$22-35
NO2	$8-25
NO3	$8-25
TN	$20-30
Bacteria (fecal coliform, fecal streptococci)	$18-25
Oil and grease	$25 -40
Pesticides (atrazine,Metolachlor)	$80-200

Unit costs for outreach and education techniques

Budgeting guidance for various outreach, neighborhood stewardship, and hotspot pollution prevention practices is provided in Tables C2-C4 (Schueler et al. 2004). These cost estimates were derived from communities across the country and should be viewed as planning level estimates. More detailed local cost analysis may be needed to get more accurate budget estimates. The costs are for implementing the techniques themselves, while additional staffing resources to plan, coordinate and administer the program are needed.

Table C2. Unit Costs for Outreach Techniques		
Technique	Unit	Estimated Cost
Overall residential outreach	Per year	$.14 - $1.11
Designer for material layout	Per hour	$100 - $150
Coloring books	Per 1,000 produced	$.45
Decals	Per 1,000 produced	$.17
Magnets	Per 1,000 produced	$.30
Posters (4 double-sided, color, 11x17)	Per 1,000 produced	$2.75
Printed materials (Flyers)	Per 1,000 produced	$.60-$.84
Printed materials (Tri-fold panel brochure)	Per 1,000 produced	$1.60 -$2.40
Stickers	Per 1,000 produced	$.08
Tote bags	Per 1,000 produced	$3.50
Billboards	Per billboard/per month	$550 -$1,850
Exterior bus advertisements	Per bus/per month	$750 - $1,450
Tabletop display	Per display	$500-$800
Educational video	Per minute of video	$1,800
Movie theatre slides	Per month	$150 -$1,400
Newspaper ads in small local paper	Per advertisement	$260 -$450
Photo displays	Per display	$121
Public attitude phone survey	Per survey of 1,000	$15,000
Radio public service announcement *	Per announcement	$40-60
TV public service announcement *	Per announcement	$2,750 - $4,000
* Assumes free airtime Sources: Council of State Governments, 1998; MacPherson and Tonning, 2003; National Oceanic and Atmospheric Administration, 1988; Water Environment Research Federation, 2000; and Center for Watershed Protection, 1998.		

Table C3. Unit Costs for Neighborhood Stewardship Practices		
Technique	**Unit**	**Estimated Cost**
Lawn care advice	Per household	$1.75 – $3.20
Rain barrel*	Per household	$20 - $45
Septic system inspections	Per household	$150-$260
Municipal Composting	Per household	$1.85 – $2.40
Soil testing	Per household	$8-$12
Compost bins	Per household	$18-$62
Curbside recycling	Per household	$29
Curbside leaf/yard waste pickup	Per household	$11.60
Household hazardous waste collection	Per household	$1.75 - $8.09
Adopt an ordinance	Per ordinance	$13,000 - $15,000
Provide stenciling materials	Per neighborhood	$300 -$400
Rain garden demonstration project	Per square foot	
	Residential	$3 to $4
	Commercial	$10 to $40
Signage	Per sign	$20-$50
"Pooper bag" stations	Per station	$250 - $300
Tree plantings	Per tree	$3.25 -$19
Pesticide advice hotline	Per year	$8,500
Non-commercial pesticide applicator licensing	Per individual	$15-$45
Cost derived from a survey of various communities across the country ** Assumes non-commercial product*		

Table C4. Unit Costs for Hotspot Pollution Prevention Practices		
Technique	**Unit**	**Estimated Cost**
Regular site inspections	Per facility	$75 - $175
Commercial lawn care/landscaping/power-washing contractors	Per individual	$25 - $75
Local ordinance to pick up non-regulated Hotspots	Per ordinance	$13,000 - $15,000
On-site illicit discharge investigations	Per facility	$220 - $900
Outreach materials to target business groups	Per hour	$30 - $45
Presentations to business groups	Per hour	$40 - $60
Non-regulatory site inspections	Per facility	$30 - $80
Business recognition programs	Per facility	$40 - $75
Discounted spill response kits, storm drain plugs, drip pans, tarps	Per facility	$60 - $250
Cost derived from a survey of various communities across the country		

References

Burton, G. and R.Pitt. 2002, Stormwater Effects Handbook. CRC Press, Boca Raton, FL.

Center for Watershed Protection. 1998. *Rapid Watershed Planning Handbook: A Comprehensive Guide for Managing Urbanizing Watersheds.* Ellicott City, MD.

Council of State Governments. 1998. *Getting in Step: A Guide to Effective Outreach in Your Community* Lexington, KY..

MacPherson, C. and B. Tonning. 2003. *Getting in Step: A Guide to Conducting Watershed Outreach Campaigns.* Prepared for the US EPA Office of Wetland, Oceans and Watersheds. EPA 841-B-03-002.

National Oceanic and Atmospheric Administration (NOAA). 1988. *Dealing with Annex V – A Guide for Ports. U.S. Department of Commerce*, NOAA, National Marine Fisheries Service, Seattle, WA. NOAA Technical Memorandum NMFS F/NWR-23 as cited in *Guidance Specifying Management Measures for Sources of Nonpoint Pollution in Coastal Waters.* EPA 840-B-92-002.

Schueler, T. R., C. Swann, T. Wright, and S. Sprinkle. 2004. *Pollution Source Control Practices – Version 1.0.* Manual 8 in the Urban Subwatershed Restoration Manual Series. Center for Watershed Protection, Ellicott City, MD.

United States Department of Agriculture. 1996. National Handbook of Water Quality Monitoring; NRCS, Washington, DC. Available at: ftp://ftp.wcc.nrcs.usda.gov/downloads/wqam/wqm1.pdf

Watershed Environment Research Federation. 2000. *Tools to Measure Source Control Program Effectiveness.* Alexandria, VA.

Appendix

D

Collection and Handling of Water Quality Samples

Most of the study designs presented in this manual involve water quality sampling techniques. Sample collection involves collecting the needed sample volume for analysis, proper non-reactive container type and preserving the sample on-site to 'stablize' for shipment and storage until it is analyzed. This Appendix provides guidance on these elements that should be incorporated into the monitoring plan for a given study.

The specific sample volume, bottle type, and preservative requirements for an individual monitoring parameter will vary and should be specified by the analytical laboratory used. *Standard Methods for the Examination of Water and Wastewater* lists the basic container requirements, minimum sample sizes, required preservative, and the maximum storage period before the analyses need to be conducted. Table D1 shows these guidelines for water samples. Care must be taken to handle the samples properly to ensure the best analytical results. Numerous losses, transformations, and changes in pollutant concentrations may occur if these guidelines are not followed. Some analyses should be conducted as soon as possible (within a few hours of sample collection, or preferably on-site or *in situ*) and include:

- CO_2
- Chlorine residual
- DO unless fixed
- Iodine
- Nitrite
- Ozone
- pH
- Temperature

Oxidation reduction potential (ORP) is also in this category of required on-site analyses, even though not in the table.

Parameters that need to be analyzed within 24-hours of sample collection (same day) include:

- Acidity
- Alkalinity
- BOD
- Cyanide
- Chromium VI (and other specific ionic forms of metals)
- Taste and odor
- Turbidity

Most of the nutrients need to be analyzed within 2 days. Many parameters can be stored for long periods of time, after preservation, specifically total forms of most heavy metals (6 months) and extracted organic compounds (30 days). In some cases, it may be possible to deviate from these guidelines if site-specific testing is conducted to demonstrate acceptable pollutant stability.

The most important guidelines are the bottle type and preservative to ensure the concentrations in the field are not changed by the time they are analyzed. Some parameters may be able to undergo longer storage periods, but this must be tested for specific conditions. The required sample volumes are all much greater than needed for most modern laboratory procedures and may be reduced if shipping costs or sample storage facilities are of a concern. Make sure that extra sample is available to redo critical analyses if problems develop as part of QA/QC procedures (see Appendix G).

Table D1. Summary of Special Sampling and Handling Requirements for Water and Wastewater Samples (*Standard Methods* 19th Edition, 1995; Collection and Preservation of Samples, 1060)[1]

Determination	Container[2]	Minimum Sample Size mL	Sample Type[3]	Preservation[4]	Maximum Storage Recommended/ Regulatory[5]
Acidity	P, G(B)	100	G	Refrigerate	24h/14d
Alkalinity	P, G	200	G	Refrigerate	24h/14d
BOD	P, G	1000	G	Refrigerate	6h/48h
Boron	P	100	G, C	None required	28d/6months
Bromide	P, G	100	G, C	None required	28d/28d
Carbon, organic, total	G	100	G, C	Analyze immediately; or refrigerate and add H_3PO_4 OR H_2SO_4 TO pH<2	7d/28d
Carbon dioxide	P, G	100	G	Analyze immediately	Stat/N.S.
COD	P, G	100	G, C	Analyze as soon as possible, or add H_2SO_4 to pH<2; refrigerate	7d/28d
Chloride	P, G	50	G, C	None required	28d
Chlorine, residual	P, G	500	G	Analyze immediately	0.5h/stat
Chlorine, dioxide	P, G	500	G	Analyze immediately	0.5 h/N.S.
Chlorophyll	P, G	500	G, C	30 d in dark	30d/N.S.
Color	P, G	500	G, C	Refrigerate	48h/48h
Conductivity	P, G	500	G, C	Refrigerate	28d/28d
Cyanide: Total	P, G	500	G, C	Add NaOH to pH>12, refrigerate in dark	24h/14d; 24h if sulfide is present
Fluoride	P	300	G, C	None required	28d/28d
Hardness	P, G	100	G, C	Add HNO_3 to pH<2	6 months/6months
Iodine	P, G	500	G, C	Analyze immediately	0.5h/N.S.
Metals, general	P(A), G(A)	500	G	For dissolved metals filter immediately, add HNO_3 to pH<2	6months/6months
Chromium VI	P(A), G(A)	300	G	Refrigerate	24h/24h
Mercury	P(A), G(A)	500	G, C	Add HNO_3 to pH<2, 4°C, refrigerate	28d/28d
Nitrogen: Ammonia	P, G	500	G, C	Analyze as soon as possible or add H_2SO_4 to pH<2, refrigerate	7d/28d

Table D1. Summary of Special Sampling and Handling Requirements for Water and Wastewater Samples (*Standard Methods* 19th Edition, 1995; Collection and Preservation of Samples, 1060)[1]

Determination	Container[2]	Minimum Sample Size mL	Sample Type[3]	Preservation[4]	Maximum Storage Recommended/ Regulatory[5]
Nitrate	P, G	100	G, C	Analyze as soon as possible or refrigerate	48h/48h (28d for chlorinated samples)
Nitrate + nitrite	P, G	200	G, C	Add H_2SO_4 to pH<2, refrigerate	None/28d
Nitrite	P, G	100	G, C	Analyze as soon as possible	None/48h
Organic, Kjeldahl	P, G	500	G, C	Refrigerate; add H_2SO_4 to pH<2	7d/28d
Oil and grease	G, wide-mouth calibrated	1000	G, C	Add HCl to pH<2, refrigerate	28d/28d
Organic compounds:		200			
MBAS	P, G	250	G, C	Refrigerate	48h
Pesticides	G(S), TFE-lined cap	1000	G, C	Refrigerate; add 1000 mg ascorbic acid/L if residual chlorine present	7d/7d until extraction 40d after extraction
Phenols	P, G	500	G, C	Refrigerate add H_2SO_4 to pH<2	*/28d
Purgeables by purge and trap	G, TFE-lined cap	2x40	G	Refrigerate; add HCl to pH<2; add 1000 mg ascorbic acid/L if residual chlorine present	7d/14d
Oxygen, dissolved: Electrode	G, BOD bottle	300	G	Analyze immediately	0.5h/stat
Winkler				Titration may be delayed after acidification	8h/8h
Ozone	G	1000	G	Analyze immediately	0.5h/N.S.
pH	P, G	50	G	Analyze immediately	2h/stat
Phosphate	G(A)	100	G	For dissolved phosphate filter immediately; refrigerate	48h/N.S.
Salinity	G, wax seal	240	G	Analyze immediately or use wax seal	6 months/N.S.
Silica	P	200	G, C	Refrigerate, do not freeze	28d/28d
Solids	P, G	200	G, C	Refrigerate	7d/2-7d
Sulfate	P, G	100	G, C	Refrigerate	28 /28d
Sulfide	P, G	100	G, C	Refrigerate; add 4 drops 2N zinc acetate/100 mL; add NaOH TO pH>9	28d/7d
Temperature	P, G	-	G	Analyze immediately	Stat/stat
Turbidity	P, G	100	G, C	Analyze same day; store in dark up to 24 h, refrigerate	24/h48h

[1] See *Standard Methods* for additional details. For determination not listed, use glass or plastic containers; preferably refrigerate during storage and analyze as soon as possible.
[2] P = plastic (polyethylene or equivalent); G = glass; G (A) or P(A) = rinsed with 1 + 1 HNO; G(B) = glass, borosilicate; G(S) = glass, rinsed with organic solvents or baked.
[3] G = grab; C = composite
[4] Refrigerate = storage at 4^o C, in the dark.
[5] Environmental Protection Agency, Rules and Regulation, 40 CFR Parts 100-149, July 1, 1992. See this citation for possible differences regarding container and preservation requirements.
N.S. = not stated in cited reference;
stat = no storage allowed; analyze immediately.

Sample Volumes

The volume of water or sediment needed is dependent on the types of toxicity assays, physical and chemical analyses, and level of precision (replicate numbers) needed. Usually one to two liters is adequate for physical and chemical analyses. The following example for determining the water volume needed for laboratory analyses is based on the requirements of a specialized in-house laboratory where shipping costs resulted in the development of analytical methods using minimal amounts of sample. It is important to work with the laboratory to determine their specific sample volume needs. Table D2 summarizes the sample quantities collected for each set of analysis. Also shown on this table is whether the sample is filtered or unfiltered (for constituent partitioning analyses). As an example, the metallic and organic toxicants are analyzed in both unfiltered and filtered sample portions in order to determine the amount of the pollutants associated with particulates and the amount that is considered "soluble." Filtering is through 0.45 μm membrane filters to determine the soluble fraction (using all-glass filtering apparatus and membrane filters that are found to have minimal affects on constituent concentrations). The sample volumes that need to be delivered to the laboratory (where further filtering, splitting, and chemical preservation will be performed) and the required containers are as follows:

- three 500 mL amber glass containers with Teflon lined screw caps
- three 500 mL HDPE (high density polyethylene) plastic containers with screw caps

A total of 3 L of each water sample is therefore needed for comprehensive analyses. In addition to the water samples, collected sediment needs to be shipped in the following sample bottles:

- one 500 mL amber glass wide mouth container with Teflon lined screw cap
- one 500 mL HDPE (high density polyethylene) wide mouth plastic container

Table D2. Example Water Volume Requirements for Different Analytes

Constituent	Volume (mL)	Filtered	Unfiltered
total solids	100 mL		Yes
dissolved solids	100 mL	yes	
turbidity	30 mL	yes	Yes
particle size (by Coulter Counter MultiSizer IIe)	20 mL		Yes
conductivity	70 mL		Yes
pH (also on-site or *in situ*)	25 mL		Yes
color	25 mL		Yes
hardness	100 mL		Yes
alkalinity	50 mL		Yes
anions (F^-, Cl^-, NO_2^-, NO_3^{2-}, SO_4^{2-}, and PO_4^{2-})	25 mL	yes	
cations (Li^+, Na^+, NH_4^+, K^+, Ca^{2+}, and Mg^{2+})	25 mL	yes	
COD	10 mL	yes	Yes
metals (Pb, Cr, Cd, Cu, and Zn)	70 mL	yes	Yes
semi-volatile compounds (by GC/MSD)	315 mL	yes	Yes
pesticides (by GC/ECD)	315 mL	yes	Yes
Microtox™ toxicity screen	10 mL	yes	Yes

The following list shows the amounts of sediment sample generally required for different chemical and physical analyses:

▪ Inorganic chemicals	90-1000 mL
▪ Organic chemicals	50-2000 mL
▪ TOC, Moisture	100-300 mL
▪ Particle size	230 - 500 mL
▪ Petroleum hydrocarbons (semi-volatile compounds)	250-1000mL

Sample Containers

Aqueous samples for toxicity testing may be collected and shipped in plastic containers. Dark borosilicate glass with Teflon® lined caps are recommended for samples to be used for organics analyses. High density polyethylene containers are needed when metals are to be analyzed. Metals can sorb to glass and also new glassware may have zinc contaminants. Polyethylene is not recommended when samples are contaminated with oil, grease, or creosote.

Wide-mouth containers made of either Teflon or high-density polyethylene, with Teflon-lined or polypropylene screw caps, are available in a variety of sizes from any scientific supply company and are considered the optimal all-purpose choice for sediment samples collected for both chemical and toxicity testing. Wide-mouth, screw-capped containers made of clear or amber borosilicate glass are also suitable for most types of analyses, with the notable exception of sediment metals, where polyethylene or Teflon is preferred. In addition, if a sediment or porewater sample is to be analyzed for organic contaminants, then amber glass bottles are recommended over plastic. It should be noted that glass containers have several disadvantages, such as greater weight and volume and susceptibility to breakage, particularly when they are filled with sediment and frozen. Plastic bags made of high-density polyethylene can also be used for storing wet or dry sediment samples for certain end uses. Generally, when the end use of the sample is known, *Standard Methods* should be consulted for specific recommendations regarding type of container, volume and storage times.

Cleaning Sample Bottles

ASTM (1996) has listed bottle cleaning/conditioning requirements in standard D 3370. New glass bottles (unless purchased pre-cleaned) must be preconditioned before use by filling with water for several days. This conditioning time can be shortened by using a dilute solution of HCl. They also point out that polyethylene is the only suitable material for sample containers when low concentrations of hardness, silica, sodium, or potassium are to be determined (in conflict with the above recommendation that warned of using polyethylene for samples containing creosote, oils or greases). All sample containers must also be sealed with Teflon™ (preferred) or aluminum lined caps. The bottles must be washed using a similar protocol as described below for the sampling equipment. ASTM (1996), in standard E 1391, also recommended more stringent preconditioning of sample containers before their first use in critical toxicological testing, as noted above (7 day leaching using a 1:1 solution of HCl and deionized water and than another 7 days in a 1:1 solution of HNO_3 in deionized water for plastics. Overnight soaking in these solutions was found to be adequate for glassware. Again, take care and test for damage before soaking equipment in strong acid solutions).

Minimum cleaning would include cleaning the samplers, including sampling lines, with domestic tap water immediately after sample retrieval. Components that can be taken to the laboratory (such as the containers in the automatic samplers) are washed using warm tap water and laboratory detergent (phosphate free), rinsed with tap, then distilled, and finally laboratory grade (18 megohm) water.

ASTM (1996) presents standard D 5088-90 covering the cleaning of sampling equipment and sample bottles. They recommend a series of washings, depending on the analyses to be performed. The first wash is with a phosphate-free detergent solution (with a scrub brush, if possible), followed by a rinse of clean (known characteristics) water, such as tap water. If inorganic analyses are to be performed (especially trace heavy metals), then the sample contacting components of the equipment and the sample bottles need to be rinsed with a 10% solution of reagent grade nitric or hydrochloric acid and deionized water. The equipment is rinsed again. If organic analyses are to be performed (especially trace organic compounds by Gas Chromatography/Mass Selective Detector), then the sample contacting components of the equipment and sample bottles need to be rinsed with pesticide grade isopropanol alcohol, acetone, or methanol. The equipment and bottles are then rinsed with deionized water and allowed to air dry. The cleaned equipment needs to be wrapped with suitable inert material (such as aluminum foil or plastic wrap) for storage and transport. If sample components cannot be reached with a brush, such as tubing, the cleaning solutions need to be recirculated through the equipment. Be careful of potentially explosive conditions when using alcohol or acetone. Work in a well-ventilated area and wear protective garments, including eye protection, when cleaning the sampling equipment with the acid or solvents.

ASTM also recommends that the equipment components that do not contact the sample be cleaned with a portable power washer or steam cleaning machine. If these are not available, then a hand brush needs to be used with the detergent solution.

Field Processing of Samples and Preparation for Shipping

If the samples are to be analyzed locally, the field collection bottles (such as the automatic sampler base with bottles) can be delivered directly to the laboratory for processing. If possible, conduct all filtering and preservation in the laboratory if at all, as this lessens the severe problems associated with field filtration and acid handling. Critical parameters (pH, DO, Eh, temperature) are analyzed *in situ* or on site. If samples cannot be delivered to the laboratory quickly, field filtration and preservation will be necessary. Samples need to be split and individually preserved, as described in *Standard Methods*. Commercial sample splitters are also availabl. A sample splitter is also useful if numerous individual sampler bottles are to be combined as a composite. The appropriate sample volumes are poured into the splitter from the individual bottles; the composite sample is then agitated and drained into individual bottles for shipping or further processing. When used to composite samples in the field and then to split into separate sample bottles, the churn must be continuously worked (about 1 to 2 seconds per stroke). Never remove the entire sample from the churn, leaving at least 20% in the bottom. Churn samplers result in sediment accumulation in the bottom of the churn, and are therefore most suitable for dissolved constituents or when larger sand-sized materials are not of interest. The use of a USGS Dekaport funnel splitter may be considered when larger particles may influence the stormwater or use the SSC method (see Box 4 in Chapter 2).

Personnel should wear latex gloves and safety glasses when handling the samples. Sample containers should be filled with no remaining head space to reduce the loss of volatile components. Samples collected for microbiological analyses or suspended solids, however, should have air-space to allow for sample mixing prior to testing.

The caps need to be screwed on securely and taped shut to reduce loosening of lid and loss of sample. The chain-of-custody seal can then be applied over the sealing tape. The paper chain-of-custody seals are not adequate to seal the lids on the jars. Do not let the water samples freeze.

Shipping Samples

Once the samples are split/divided into the appropriate shipping bottles (and preserved, if needed), the sample container label should be filled out completely and then logged onto a shipping list for each shipping container. Shipping containers are usually plastic coolers. There needs to be adequate packing (preferably as many "ice" packs as can fit, plus bubble wrap) inside the shipping container to insure that the sample bottles do not rub or bang against each other en route. Do not use packing peanuts (especially the water soluble type) to fill up space. Wrap glass bottles with bubble wrap. Use sufficient "blue ice" or other cooling packs to insure the coolers stay cool during shipment. Do not use water ice. The coolers must also be securely taped shut (seal the seams) to minimize leakage if a bottle breaks during shipment. The samples should be sent via overnight courier and timed so they arrive while laboratory personnel are present and sufficient time is available to initiate the critical analyses immediately (unless special arrangements have been made with the laboratory). Always call to schedule a sample shipment and fax a confirmation of the sample shipping information. Always keep a copy of any sample identification sheets and send the originals (by mail, not in the coolers). Include a shipping list (and copy of appropriate sampling forms) in an envelope taped to the outside of the cooler.

References

ASTM (American Society of Testing and Materials). 1996 *Annual Book of ASTM Standards*. West Conshohocken, PA: ASTM, vol. 04.08, 1996.

Standard Methods for the Examination of Water and Wastewater. 19th edition. Water Environment Federation. Washington, D.C. 1995.

Appendix

E

Social Survey Methods: Estimating sample size and survey errors

Estimating Sample Size

In determining the goals of the survey, one should account for size of the population, the actual data being sought, and the resources available for surveying. The level of precision refers to the sampling error (for example, 3%), the range of which is called the margin of error (for example, 43-49%). The combination of resources and desired accuracy will dictate the acceptable margin of error. Another factor is the desired confidence level. The most common confidence level is 95%, meaning that in a survey of 100 individuals, 95 could be considered accurate values at the *precision level* indicated previously. The larger the sample size, the higher the confidence level attained. The degree of variability relates to the heterogeneity of the population- if the population is very mixed in relation to the concept being surveyed, the sample size must be larger; if they are more divided, the sample size can be smaller (while achieving a similar confidence level). Finally, in the case of social surveys, not all who are sampled will reply, and the rate of response must be taken into account. Using past data on the response rate of similar surveys, divide the expected percentage of response by the desired sample size and arrive at the actual size required for survey.

Table E1 provides a table of values to estimate sample size based on a confidence level of 95% and margins of error for 5% and 10%. The level of precision is indicative of sample size alone and does not reflect any other types of errors that may arise to reduce the level of precision of the survey results.

Table E1. Guidance to select sample sizes for two different precision levels (from Taylor-Powell 1998)

Population Size	Sample Size		Population Size		Sample size	
	5%	10%			5%	10%
10	10		275		163	74
15	14		300		172	76
20	19		325		180	77
25	24		350		187	78
30	28		375		194	80
35	32		400		201	81
40	36		425		207	82
45	40		450		212	82
50	44		475		218	83
55	48		500		222	83
60	52		1000		286	91
65	56		2000		333	95
70	59		3000		353	97
75	63		4000		364	98
80	66		5000		370	98
85	70		6000		375	98
90	73		7000		378	99
95	76		8000		381	99
100	81	51	9000		383	99
125	96	56	10000		385	99
150	110	61	15000		390	99
175	122	64	20000		392	100
200	134	67	25000		394	100
225	144	70	50000		397	100
250	154	72	100000		398	100

Online sample size calculators are available:
http://www.surveysystem.com/sscalc.htm
http://www.gifted.uconn.edu/siegle/research/Samples/samplecalculator.htm

Survey Sampling Errors

The effectiveness of any survey is in large part dependent upon how well the survey is designed and how the survey sample is selected. A sample is intended to estimate the characteristics of the population so that generalizations to the larger population may be made. These generalizations may not be possible if sampling errors occur. There are two major types of errors: sampling and non-sampling. Sampling errors result from errors with the methods used to identify the sample population. Adherence to scientific methods is needed to ensure the sample is properly identified, and in part, a sufficient sample size is selected. If a random, or probability-based, method is used to select a sample from a population, the list of all possible participants to choose from should be:

- accurate, including only those individuals of interest;
- complete and current, including all individuals of interest;
- free of duplicate names; and
- absent any patterns in the way the names are listed (Taylor-Powell 1998).

If the contact information for the population is not accurate, this will increase the non-response rate for the survey and lead to less accurate findings. Screening the list for any bias or pattern in the ordering of names or places is needed.

Other types of errors that may affect the accuracy of the survey results are referred to as 'non-sampling' or measurement error and occur when, for example, respondents provide an incorrect answer, or a survey is not completed (e.g. a non-response). Analysis of the pilot (or pre-test) survey results should identify and correct the issues that may arise with the type of survey questions, how they are worded and the types of responses provided. For incorrect, or unmatched answers where a respondent provides an answer different than those provided on the survey, Scheurin (2004) suggests entering the average response.

Appendix

F

Comparisons of Regulatory Agency Monitoring Protocols for Structural and Non-Structural Stormwater Treatment Practices

A summary of two state-level monitoring protocols used to evaluate stormwater treatment practices, to include proprietary devices, is provided in this Appendix. (the Technology Acceptance Reciprocity Partnership (TARP) and Washington State's Technology Assessment Protocol – Ecology (TAPE)). The TARP protocol is used to ensure that stormwater treatment practices are evaluated in a uniform manner assuring minimum standards for quality assurance and quality control (QA/QC). The TAPE protocol is used to characterize, with a reasonable level of statistical confidence, the pollutant removal effectiveness of emerging stormwater treatment technologies for an intended application and to compare test results with vendor's claims. CWP (2008) provides a more complete review of the scope, purpose and evaluation procedures used by these and other monitoring protocols for proprietary devices.

Feature	TAPE		TARP	
	Explanation	Criteria	Explanation	Criteria
States Applicable	States apply this requirement	Washington	States apply this requirement	California, Massachusetts, Maryland, New Jersey, Pennsylvania, Virginia
Number of events, minimum per application (#)	Number of events used to evaluate device and may also depend on statistical evaluation methods	12-35	At least 50% of the total annual rainfall must be sampled, for a maximum of 15 inches of precipitation	15-20
Composite sampling methods	Required composition method	Flow-weighted and discrete flow rate composite sampling	Time-weighted composite samples are not acceptable, unless flow is monitored and the event mean concentration can be calculated from the data	Flow-weighted composite sampling which follows the NPDES guidance
Minimum storm depth (inch)	Total rainfall amount during the sampling event	0.15	More than 0.1 inch of total rainfall	0.1
Storm start/end (antecedent dry period)	Defines the storm event's beginning and end as designated by minimum time interval without significant rainfall	6 hours minimum less than 0.04 inches of rain	A minimum inter-event period	6 hours
Minimum storm duration (hr)	Shortest acceptable runoff duration	1	N.A.	N.A.
Minimum storm intensity	Lowest intensity that qualifies as a rainfall event	None, as long as above criteria are met	N.A.	N.A.
Sampling methods	To use automatic sampling equipment for insoluble TPH/oil, a determination is needed, supported by appropriate data, that any TPH/oil adherence to the sampling equipment is accounted for and meets QA/QC objectives	Automatic samplers, except for chemical constituents that require manual grab samples	Grab samples should only be used for certain constituents	Programmable automatic flow samplers with continuous flow measurements
Sampling for organic contaminants	Required tube	Tygon or Teflon Tube	Required tube	Tygon or Teflon Tube
Flow-weighted composite sampling	Samples are collected over the storm event duration and composited in proportion to flow	10 aliquots should be composited, covering at least 75% of each storm's total runoff volume up to the design storm volume	A minimum of 10 water quality samples (10 influent and 10 effluent) should be collected per event. For composite samples, a minimum of 5 subsamples is acceptable	Obtain flow-weighted composite samples covering a minimum of 70% of the total storm flow, including as much of the first 20% of the storm as possible

Feature	TAPE		TARP	
	Explanation	Criteria	Explanation	Criteria
Discrete flow composite sampling	Samples representing relatively constant flow period (less than 20% variation)	80% TSS removal at the design hydraulic loading rate	N.A.	N.A.
Sampling location	Provide a site map showing all monitoring/sampling station locations and identify the equipment to be installed	Samples should be collected at a location where the stormwater flow is well-mixed	Provide a site map showing all monitoring/sampling station locations and identify the equipment to be installed	Influent location will be directly upstream of the system and before the flow is split between the treatment system and the bypass. Effluent sampling location will be directly downstream of the treated flow and after the effluent joins the bypass
Flow monitoring interval	Flow into and out of the treatment device must be measured and recorded on a continuous basis over the sampling event duration	15 min or shorter interval	N.A.	N.A.
Rainfall monitoring interval	Gauge should be calibrated at least twice during the field test period	15 min or shorter interval	15 min increments are recommended for consistency with NWS reporting	15 min
Target pollutants	One or more of following parameters	TSS, nutrients, heavy metals (cadmium, copper, lead, zinc, phosphorus), petroleum hydrocarbons, and toxicity	In selecting test parameters, include TSS and SSC at a minimum, and consider other parameters	TDS, TSS, SSC, TPH, TKN, total (nitrogen, phosphorus, COD, BOD), E-coli, Total Coliform, enterococci, pH, conductivity, temperature, lead, copper, zinc, and nickel
Sampling for Total Suspended Solids (TSS)	Particles greater than 500 microns are being considered similarly to debris and other gross solids	80 percent TSS removal for basic treatment level	N.A.	N.A.
Particle size distribution (PSD)	TSS and PSD should be analyzed prior to installing the treatment device	80 percent TSS removal for basic treatment level	N.A.	N.A.
Accumulated sediment sampling procedure	Following sediment constituents should be analyzed. The sediment sample should be a composite from several grab samples (at least four) collected from various locations within the treatment system	Percent total solids, grain size, total volatile solids, NWTPH-Dx	N.A.	N.A.

Feature	TAPE		TARP	
	Explanation	Criteria	Explanation	Criteria
Sample preservation, handling, and analysis	Samples are to be preserved and analyzed in accordance with following methods	Washington State Department of Ecology methods, US EPA-approved methods, or Standard Methods	Samples are to be preserved and analyzed in accordance with following methods	ASTM Methods as listed in Appendix B of TARP and Certified Laboratory
Quality control	Small batches (less than 20 samples) should include one of each type of quality control procedure specified	QC should be performed on no less than 10 percent of the analyzed samples.	N.A.	N.A.
Laboratory tests	Laboratory tests should be conducted under the following conditions	Constant flow rates of 75, 100, and 125 percent, plus or minus 10%; U.S. Silica Sil-Co-Sil 106 ground silica can be used; 100 and 200 mg/L TSS influent concentration range	N.A.	N.A.

Appendix

G

Elements of a QA/QC Plan

A Quality Assurance/ Quality Control (QA/QC) plan is needed to provide confidence in the data and results generated as part of the monitoring study. The QA/QC plan describes the sample collection and laboratory analyses procedures that will be followed in order to limit the errors in sampling or analyzing the data. A project team member should be identified to be responsible for the oversight of QA/QC development and implementation. This Appendix is not intended to provide a comprehensive review of QA/QC techniques but to familiarize communities with an overview of important elements of a QA/QC plan to include in the development of a monitoring program.

Field QA/QC

Field QA/QC procedures begin before the first sample is taken through receipt of the samples to the laboratory for analysis. A field form similar to Figure G-1 records the date, time, sample type and identification number along with field conditions at the time of samples. These field forms are kept in a project binder or entered into an electronic database. This information helps to 'redflag' unusual conditions or samples than may need further evaluation. QA/QC procedures for field sampling include:

- Define storms eligible for sampling
- Sample collection procedures and transport QA/QC
- Equipment decontamination
- Appropriate field sample containers and labeling
- Chain-of-custody records completed
- Sample receipt

Storms selected for sampling

Defining the storms that are eligible for sampling requires some understanding of the local or regional rainfall patterns and frequency that can be achieved by a frequency analysis of long term rainfall records. Example criteria include:

- No greater than a trace of precipitation 24hrs preceding the storm event
- Must be a runoff generating rainfall
- Runoff must be sufficient to generate the require water volume for constituent analysis

Careful consideration is needed to ensure the criteria are not too stringent or there are too many criteria that are difficult to collectively meet and results in few samples collected for analysis. For example, using criteria for storm selection in the State of Washington resulted in only 4 storm events meeting the criteria out of 21 – too many or stringent criteria to satisfy (Lenhart 2007). This had a sampling cost of $84,000 ($4000 per storm for 21 samples), but only 4 of the samples were 'keepers'.

<table>
<tr><td>Date:</td><td colspan="4">Time:</td><td>Sample collector(s)</td><td colspan="2"></td></tr>
<tr><td>Sample location:
(i.e., street, block number, closest intersection)</td><td colspan="4"></td><td rowspan="2">Purpose
(circle)</td><td rowspan="2" colspan="2">Stormflow

Baseflow</td></tr>
<tr><td>Type:
(circle all that apply:</td><td>Water flow</td><td>Water quality</td><td>Sediment</td><td>Particulate</td></tr>
<tr><td>Number of samples:</td><td></td><td></td><td></td><td></td><td>Precipitation (past 24 hours)</td><td>Yes</td><td>No</td></tr>
<tr><td>Sample i.d.
(same as custody form)</td><td colspan="4"></td><td>Precipitation amount, if applicable (inches)</td><td></td><td>n/a</td></tr>
<tr><td>Chain of Custody identification number</td><td colspan="4"></td><td>Data Source for precipitation</td><td colspan="2"></td></tr>
<tr><td rowspan="2">Lab(s) for analyses
(circle)</td><td rowspan="2" colspan="4">Lab 1
Lab 2
Lab 3</td><td>Temperature (F)</td><td colspan="2"></td></tr>
<tr><td>Wind speed</td><td colspan="2"></td></tr>
</table>

Figure G1. Example field log sheet for sample collection

Sample collection and transport QA/QC

It is important to understand the methods used to collect the sample as this will impact the analytical methods themselves to characterize pollutant concentrations. The correct type of sample containers need to be used and the containers need to be filled according to sample volume requirements where water quality samples will be filled to the top minimizing head-space. All samples collected in the field should be placed in a cooler maintained at or below 4 ºC for transport to the appropriate analytical laboratory. Collected samples will be transported to the laboratory the same day and stored in a refrigerator at 4 ºC until analysis. Appendix D summarizes information on requirements for sample volume, bottle type, preservative and shipping requirements for water quality samples.

Equipment decontamination

Throughout sample collection, care is needed to avoid sample contamination. It is important that field personnel complete their training and understand the data-quality requirements (e.g. sample bottles, holding times) and potential sources of contamination (USGS 2006). This is accomplished through rigorous decontamination procedures and careful sample handling procedures, including

- Sampling equipment will be cleaned prior to use and between samples
- All sources of sample contamination (airborne sources, fingers, unclean equipment) should be avoided
- New, pre-cleaned sample bottles will be used for every sample collection
- Always wear gloves to collect sample and avoid touching rim of container

It is suggested that the USGS (2006) "clean hands dirty hands" approach to sampling be used. This technique specifies duties performed by field personnel with 'clean' hands to collect and prepare samples for transport while the other 'dirty' hands are not in contact with the samples and are involved in equipment maintenance etc.

Field Sample Containers and Labeling

Preprinted labels that include a unique sample identification number are needed for each sample container (Figure G2). The sample ID prefix may reflect the monitoring data or site being collected. This provides a unique link between the field records and each sample. Sample collection information is hand-recorded in bound, pre-paginated logbooks, then keyed into spreadsheets or project-specific applications. Data entry into the electronic format follows the sampling efforts.

Outfall monitoring SB-101
Unique Sample ID: XX-______
Station No.______________________________
Sample Type (circle one): water/ sediment
Date: __________ Time: ___________
Sample Collector:__________________
Container____of______
GPS coord. ________________

Figure G2. Example of a label for a sample bottle

Chain of Custody

Sample chain of custody (COC) records are the administrative records associated with the physical possession and/or storage history of each individual sample from the purchase and preparation of each sample container and sampling apparatus to the final analytical result and sample disposal. An example custody record is provided at the end of this Appendix. Sample custody is documented throughout collection, shipping, analysis, and disposal of the sample. Samples should not be left unattended unless properly secured. A separate COC record must accompany each cooler brought to the laboratory. In addition, the outside of the coolers are marked to indicate the number of coolers in the shipment (e.g., 1 of 2, 2 of 2). Each analytical laboratory should maintain a formal, documented system designed to provide sufficient information to reconstruct the history of each sample, including preparation of sampling containers, sample collection and shipment, receipt, distribution, analysis, storage or disposal, and data reporting within the laboratory. Laboratory documentation must provide a record of custody for each sample (versus a sample batch) throughout processing, analysis, and disposal.

The COC form summarizes the samples collected and analyses requested and tracks sample release from the field to the initial receiving laboratory. An example COC form used for a Chesapeake Bay Street Sweeping study is provided at the end of this Appendix (Figure G3). Each sample custody form needs to be signed by the person relinquishing samples once that person has verified that the custody form is accurate (i.e., that all samples present in the shipping container are listed on the form, and that the sample descriptions, requested analytical methods, and sampling dates are accurate). The original sample custody forms accompany the samples, and the person delivering the sample keeps a copy. Upon receipt at the sample destination, sample custody forms are signed by the person receiving the samples once that person has verified that all samples identified on the custody forms are present in the shipping container. Any discrepancies are noted on the form (in addition to any internal laboratory documentation policy) and the sample receiver should immediately contact the Project Manager to report missing, broken, or compromised samples. Samples are considered to be in a person's custody (custodian) if:

- The samples are in a person's actual possession
- The samples are in a person's view after being in that person's possession
- The samples were in a person's possession and then were locked or sealed up to prevent tampering, or
- The samples are in a secure area

Sample Receipt

Immediately upon receipt by a laboratory, the condition of samples must be assessed and documented. The contents of the shipping container must be checked against the information on the custody form for anomalies. If any discrepancies are noted, or if laboratory acceptance criteria or project-specific criteria are not met, the laboratory must contact the Project Manager for resolution of the problem. The discrepancy, its resolution, and the identity of the person contacted must be documented in the project file. The following conditions may cause sample data to be unusable and must be communicated to the laboratory team manager:

- The integrity of the samples is compromised (e.g., leaks, cracks, grossly contaminated, container exteriors or shipping cooler interiors, obvious odors, etc.)
- The identity of the container cannot be verified
- The proper preservation of the container cannot be established
- Incomplete sample custody forms (e.g., the sample collector is not documented or the custody forms are not signed and dated by the person who relinquished the samples)
- The sample collector did not relinquish the samples or,
- Required sample temperatures were not maintained during transport.

The sample custodian must verify that sample conditions, amounts, and containers meet the requirements for the sample and matrix (e.g. water, sediment or soil). A unique sample identifier must be assigned to each sample container received at the laboratory, including multiple containers of the same sample. This should correspond with the sample label.

Laboratory QA/QC

Laboratory analysis of samples is one of the most expensive parts of any monitoring study. Any success with getting a sufficient number of samples in the field may be jeopardized by poor laboratory analysis, therefore QA/QC is very important. During the scoping and budgeting process, program staff needs to decide who will do the analyses and ensure proper QA/QC procedures are followed, whether in-house or through private contract facilities. A certified or accredited laboratory (in-house or contract) will have QA/QC procedures prepared that should be reviewed and the necessary elements incorporated into the overall monitoring plan. Some considerations for selecting and working with a contract lab include:

- Make sure that the lab is EPA-certified for the indicator parameters chosen. A state-by-state list of EPA certified labs for drinking water can be found here: http://www.epa.gov/safewater/privatewells/labs.html. State environmental agencies are also good resources to contact for pre-approved laboratories. Another measure of technical competency for laboratories is accreditation by one of the five major accrediting bodies in the U.S. Laboratory accreditation is based on internationally accepted criteria for competence (ISO/IEC 17025:2005).
- Make sure the operators have up-to-date certification. Adequate training and suitable experience of analysts are necessary for good laboratory work. Periodic tests of analytical skill are needed.
- Choose a lab with a short turn-around time. As a rule, a lab should be able to produce results within 48 hours.
- Clearly specify the indicator parameter and analysis method desired, using the guidance in this manual or advice from a water quality expert. EPA analytical guidelines published in the *Federal Register* for the different tests specify the types and magnitude of QA/QC analyses. Table G1 lists an example table from a QA/QC plan for analyzing street particulate matter for a street sweeping study (from CWP 2006).
- Ensure that the maximum hold time (the time between when the sample is collected and analyzed) for each indicator parameter exceeds the time it takes to ship samples to the lab for analysis.
- Look for labs that offer electronic reporting of sample results, which can greatly increase turn-around time, make data analysis easier, and improve response times.
- It is recommended that the same consultant be used throughout the course of any particular monitoring study, if at all possible. Small differences in lab procedures and equipment may render data from different consultants incomparable. Certain analysis methods have some subjectivity that will affect the results depending on who performs the analysis. For example, with macroinvertebrate identification, taxonomic ambiguities may arise when results are reported at multiple taxonomic levels. The method used to resolve ambiguous taxa can strongly influence the analysis and interpretations of assemblage data (Cuffney et al. 2005). This will also include identifying the specific chemical species to analyze for a water quality parameter such as total, dissolved, or particulate if using two different labs to verify analytical results using split samples.

Table G1. Typical List of Standard and Modified Methods for Wet Weather Flow Analyses	
Parameter	**Method**
Physical Analyses	
Color, Spectrophotometric	EPA 110.3
Conductance, Specific Conductance	EPA 120.1
Particle size analysis by Coulter Counter and sieves	Coulter method
pH, Electrometric	EPA 150.1
Residue, filterable, gravimetric, dried at 180 ^{o}C	EPA 160.1
Residue, non-filterable, gravimetric, dried at 103-105 ^{o}C	EPA 160.2
Residue, total, gravimetric, dried at 103-105 ^{o}C	EPA 160.3
Residue, volatile, gravimetric, ignition at 550 ^{o}C	EPA 160.4
Turbidity, nephelometric	EPA 180.1
Inorganic Analyses	
Hardness, Total (mg/L as $CaCO_3$), Titrimetric EDTA	EPA 130.2
Aluminum, arsenic, cadmium, chromium, copper, iron, lead, nickel, and zinc	EPA 200.9
Chloride, fluoride, nitrate, nitrite, phosphate, and sulfate	EPA 300.0
Ammonium, calcium, lithium, magnesium, potassium, and sodium	EPA 300.0 modified
Alkalinity, titrimetric (pH 4.5)	EPA 310.1
Organic Analyses	
Chemical Oxygen Demand, colorimetric	EPA 410.4
Aldrin, Chlordane-alpha, Chlordane-gamma, 4,4'-DDD, 4,4'-DDE, 4,4'-DDT, Dieldrin, Endosulfan I, Endosulfan II, Endosulfan sulfate, Endrin, Endrin aldehyde, Endrin ketone, HCH-alpha, HCH-beta, HCH-gamma (Lindane), Heptachlor, Heptachlor epoxide, and Methoxychlor	EPA 608 modified
Acenaphthene, Acenaphthylene, Anthracene, Azobenzene, Benzo(a)anthracene, Benzo(b)fluoranthene, Benzo(g,h,i)perylene, Benzo(k)fluoranthene, Benzo(a)pyrene, 4-Bromophenyl-phenylether, Bis-(2-chloroethyl)ether, Bis-(2-chloroethoxy)methane, Bis-(2-ethylhexyl)phthalate, Butylbenzyl phthalate, Carbazole, 4-Chloro-3-methylphenol, 2-Chloronaphthalene, 2-Chlorophenol, 4-Chlorophenyl-phenylether, Chrysene, Coprostanol, Dibenzo(a,h)anthracene, 1,2-Dichlorobenzene, 1,3-Dichlorobenzene, 1,4-Dichlorobenzene, 2,4-Dichlorophenol, Diethyl phthalate, 2,4-Dimethylphenol, Dimethyl phthalate, Di-n-butyl phthalate, 2,4-Dinitrophenol, 2,4-Dinitrotoluene, 2,6-Dinitrotoluene, Di-n-octyl phthalate, Fluoranthene, Fluorene, Hexachlorobenzene, Hexachlorobutadiene, Hexachlorocyclopentadiene, Hexachloroethane, Indeno(1,2,3-cd)pyrene, Isophorone, 2-Methylnaphthalene, 2-Methylphenol, 4-Methylphenol, Naphthalene, Nitrobenzene, 2-Nitrophenol, 4-Nitrophenol, N-Nitroso-di-n-propylamine, N-Nitroso-diphenylamine, Pentachlorophenol, Phenanthrene, Phenol, Pyrene, 1,2,4-Trichlorobenzene, 2,4,Trichlorophenol, and 2,4,6-Trichlorophenol	EPA 625 modified
Toxicity Analyses	
Microtox™ 100% toxicity screening analysis (using reagent salt for osmotic adjustments)	Azur Environmenal method

Review the lab's QA/QC procedures (discussion with water quality specialist/analyst) to ensure the sampling equipment is functioning properly and provided accurate results. The procedures for cleaning equipment and calibrating instruments should also be evaluated. These QA/QC procedures are described below.

- *Lab spikes* – Samples of known concentration are prepared in the laboratory to determine the accuracy of instrument readings.

- *Sample blanks* are used to ensure the analysis equipment is operating properly, solvents/reagents used in analysis are free of impurities or are of known concentrations, and the operator is using the equipment properly. Some examples include:
  - *Instrument/Equipment blanks* – Used to establish baseline response of an instrument in the absence of an analyte (substance being analyzed), to detect contamination with the sampling equipment and to verify the effectiveness of cleaning the sampling equipment.
  - *Calibration blank*s - (solvent blank). Used to detect and measure solvent (a substance that dissolves another substance or substances to form a solution) impurities. Similar to the above blank but only contains the solvent used to dilute the sample. This typically is the "zero" concentration in a calibration series.
  - *Method blank* (reagent blank). Used to detect and measure contamination from all of the reagents (substance or compound consumed during a chemical reaction) used in sample preparation. A blank sample (using ultrapure water) with all reagents needed in sample preparation is processed and analyzed. This value is commonly subtracted from the analytical results for the samples prepared in the same way during the same analytical run.
  - *Trip blank* (sampling media blank). Used to detect contamination associated with field filtration apparatus and sample bottles. A known water (similar to sample) is carried from the laboratory and processed in the field in an identical manner as a sample.
  - *Split samples* – Samples are divided into two separate samples at the laboratory for a comparative analysis. Any difference between the two sample results suggests the analysis method may not be repeatable. Split samples may also be used to determine if analysis can detect a known addition to a sample.
  - *Analysis of duplicates* – Standard Methods (date) suggests that at least 5% of the samples have duplicate analyses.
- *Equipment cleaning and instrument maintenance protocols* – Each lab should have specific and routine procedures to maintain equipment and clean glassware and tubing. These procedures should be clearly labeled on each piece of equipment.
- *Instrument calibration* – Depending on the method, instruments may come with a standard calibration curve, or may require calibration at each use. Lab analysts should periodically test the default calibration curve. Consult the laboratory to determine what calibration procedures are used (e.g. check for the use of control charts and comparison to standards).

Once the data is analyzed and returned to the project team, the final step is data management where specific procedures are defined to 'enumerate' non-detect values. Non-detect values of samples are samples whose concentrations or other measurement value are below the detection limit of the analytical methods. A detection limit is the smallest value that can be detected above the background noise using a specific procedure with a specific confidence. There are several different methods used to determine detection limits used in laboratory analyses (e.g. instrument detection limit, method detection limit, practical quantification limit). Non-detect values do not present a serious problem if they present a small fraction of the dataset. However, if the detection limit available results in many non-detected data (say between 25 and 75% of the observations), statistical analyses are severely limited. Data records used for analyses need to indicate by using a qualifier next to the value that identifies the value as an estimate of the true value. There are a variety of approaches for handling non-detect values to avoid omitting these data from the analyses. Common 'replacement' methods include using half of the detection limit, using zero, or the detection limit. More robust statistical methods such as regression on ordered statistics (e.g. Helsel 2005) may also be used.

Once the analytical results are available, the data are reviewed internally by the laboratory, in addition to program staff once the data are received. The laboratory reviews the data and compares them to the data quality objectives Data review should also check for extreme values that should be questioned but not necessarily discarded as "outliers." In all cases, unusual analytical results should be reported to the field sampling crew and other personnel as soon as possible to solicit their assistance in verifying that the results are valid and not associated with labeling or sampling error.

CHAIN OF CUSTODY SHEET AND SAMPLE REQUEST

Chain of Custody No:	Sample collector:	Date Samples Despatched:	Sheet ………………
EPA project: Effectiveness of Street Sweeping/Storm Drain Cleanout	Address:	Needs refrigeration at or below 4 oC	of ……………………
		Sampler :	
	Tel: Fax:		
	Project/Site Name:		
	Email :		
	Contact Name:		

Date of Sampling:	**Sample Ref. ID**	**Sample location**	Sample Preservation Y/N	(S)oil or (W)ater (specify if other)	**Suite Name/Analysis Required**	**Turnaround - please tick** 14 day t/a ____ 7 day t/a ____ 4 day t/a____ 3 day t/a ____ other ____ (please specify) or date results required by : **Comments**	**Report format** Standard___	**Samplers Signature**

Special Instructions:

Date Received: *Time:* *Signature:*

Figure G3. Example chain of custody form.