BASMAA Alternative Trash Assessment Methodology
On-Land Trash Cleanup Monitoring Study
Final Study Design Concept
August 17, 2016

1.0 On-Land Trash Cleanup Monitoring Study Objectives

The main goal of performing on-land cleanups is to remove as much trash as possible to reduce loads to the stormwater drainage system. The main goal of the monitoring study is to develop a protocol to demonstrate trash load reduction effectiveness associated with on-land cleanups. The monitoring study aims to rigorously quantify trash load reductions associated with on-land cleanups and demonstrate reliable progress towards changes in trash generation rate categories (i.e., a color change on trash generation maps). The monitoring study will test different frequencies of on-land cleanups, performed using a prescribed protocol, in moderate, high, and very high trash generating areas. Trash levels before and after on-land cleanups will be measured using on-land visual assessments and through quantitative measurement of (1) trash picked up during cleanup activities in the study area and (2) trash contained in full trash capture units draining the study area.

The primary objective of quantifying trash that is picked up during cleanup activities is to identify how much trash is removed by the activity; a secondary objective is to validate the trash generation category that is present prior to the cleanup event occurring. The study will also measure the volume of trash present in drain inlets (DIs) equipped with full trash capture devices, which is consistent with how trash generation rates were originally established via the BASMAA Trash Generation Study. The amount of trash present in the full trash capture device represents the trash generation rate/category that results from the cleanup activity. How the measurements will be used is illustrated in the load reduction calculations provided in Section 5.0.

The trash reduction from on-land cleanups in areas with different baseline trash generation categories (e.g., moderate, high and very high) and different cleanup frequencies (e.g., monthly, weekly, daily) will be studied. On-land cleanups may be performed by volunteers and/or contractors depending on the municipality but all will adhere to the prescribed protocol. There will be an attempt to select monitoring sites that represent a variety of sources and types of trash.[1]

The desired outcome of the monitoring study is the development of a Performance Standard[2] and specified management action that includes:

(1) a standardized on-land cleanup protocol and;

(2) trash reduction values (percent reduction) related to a given baseline trash generation category and cleanup frequency.

[1] The Water Board may review proposed monitoring study sites for representativeness.
[2] MRP Provision C.10.b.ii.b(iv) states "Permittees may put forth substantive and credible evidence that certain management actions or sets of management actions when performed to a specified performance standard yield a certain trash reduction outcome reliably".

2.0 Proposed Study Design Parameters

The following are the proposed key study design parameters (summarized in Table 1, below):

- **Number of baseline trash generation categories tested:** 3 (B [moderate], C [high], D [very high])
- **Number of on-land clean up frequencies tested**: 3 (e.g., daily[3], weekly, monthly[4]). The final cleanup frequencies that will be selected depend on the specific cleanup activities of participating municipalities.
- **Definition of "site" area**: A study "site" is defined as the drainage area for one to two DIs that contain full trash capture devices. Cleanup activities must occur within the entire drainage area to the DIs.
- **Number of sites**: 3 sites for each baseline trash generation category and cleanup frequency combination: 8 combinations * 3 sites = 24 "study sites".
- **Number of monitoring events for each study site**: 10. Monitoring events will cover the wet season and the dry season with the majority of events covering the wet season (exact dates to be determined). The study will also focus on prioritizing monitoring at the end of the dry season. The number of monitoring events is summarized in Table 1.
  - The monitoring frequency/ no. of events sampled could potentially be increased for sites with daily cleanup activities, to better capture the daily effect.
  - The number of monitoring events needed to produce results within an acceptable degree of confidence depends on the variability of the data. Based on power analyses conducted by EOA and presented to Water Board staff on May 12, 2016, on average a minimum of 3 visual assessments are needed at a site to determine (with confidence) a change in one trash generation category (see Attachment #1). On average 10 assessments are needed to detect a change of ½ a trash generation category at a site. Given these results, the Project Team is comfortable recommending that a total of 10 assessments/events occur at each study site.

[3] A "daily" cleanup frequency= 5-7 times per week.
[4] Areas with very high trash generation rates are not proposed to be tested at a monthly frequency, because reduction from monthly cleanups would likely not be measurable in these areas.

Table 1: Proposed No. of Study Sites and Monitoring Events

Proposed No. Study Sites				**Total No. Sites**	**Total No. Monitoring Events per Site**	**Total No. Monitoring Events**
Trash Generation Rate	*Cleanup Frequency*					
	Daily	Weekly	Monthly			
Medium	3	3	3	9		
High	3	3	3	9		
Very High	3	3	0	6		
				24	10	240

Notes:

1. The majority of monitoring events is proposed to occur during the wet season (e.g., ≥75 percent).
2. The number of monitoring events may be increased for study sites with a daily cleanup frequency.

- **Timing of Visual Assessments:** Visual Trash Assessments will be performed immediately before and after the cleanup activity. The before cleanup assessment will be used to assess the trash generation rate that is present prior to the cleanup event occurring. The post clean-up assessment will be used to validate that cleanup activities followed the prescribed protocol and the site was cleaned to an "A" score. Analysis of the extensive assessments conducted by EOA to-date have demonstrated that temporal variability (assessments performed at different times of the year) does not appear to have a significant effect on baseline assessment scores (see Attachment #1).
- **Other data collection efforts.** Field data sheets will be developed to facilitate the collection of ancillary data that could help explain potential outliers in the study results. Examples include a one-time occurrence of illegal dumping, or the occurrence of a street festival prior to a monitoring event. In addition, it will be important for the participating municipalities to inform the monitoring team about any out-of-the-ordinary trash-related factors occurring in-between monitoring events during the monitoring period.
- **Full trash capture device maintenance schedules:** The degree of vegetative debris loading affects both the frequency and timing of device maintenance. This variable will be considered in the monitoring study design, and the specific (historic) maintenance schedule for each study site selected will be included in the SAP and accounted for in the schedule of monitoring events. Trash capture device maintenance performed during the study period will be documented and the amount of trash versus debris removed will be used in assessing the results.
- **Use of a "Control" site** (i.e., testing trash loading where no on-land cleanup activities are occurring): The study will not test "control" sites. Control sites are not needed based on how the load reduction is proposed to be calculated (see Section 5.0). However, the study may draw from data collected as part of the BASMAA Tracking California's Trash Study to evaluate trash loading in areas where on-land cleanups are currently not occurring.

- **Performance Standard Applicability & Variables.** The performance standard will describe how and where the cleanup protocol may be applied. For example, if the performance standard was applied to evaluate the trash load reduction in an entire Trash Management Area where there is a significant amount of private property draining to DIs (if private property is not cleaned as part of the cleanup activity), the percent reduction must be weighted based on the percentage of private property draining to the MS4. Other factors that affect the applicability of the Performance Standard as learned through the monitoring study implementation will be incorporated into the Performance Standard. Other factors may include but are not limited to the street sweeping schedule, whether or not on-land cleanups are conducted during rain events, and when cleanups should be performed with respect to predicted large rain events.

3.0 Site Selection Criteria

The following are the proposed site selection criteria:

- Meets designated trash generation rate for study site.
- The level of trash in the area is consistently available and transportable to the MS4 (e.g., this does not preclude sites where illegal dumping occurs provided the activity occurs on a consistent basis).
- Consistent, known cleanup frequency across designated sites, to be monitored to ensure proper cleanup protocols are implemented.
- Includes a minimum of one drain inlet (DI). All DIs in a designated study area will have a full trash capture device installed. Auto retractable screens should not be installed, as they will prevent some trash from entering the drain inlet, where trash will be collected and quantified.
- The on-land cleanup activity covers the entire drainage area of the DI(s). This precludes, for example, drainage areas comprised of private property, where cleanup activities are not being performed by the municipality. The drainage area for the DI will need to be delineated for the SAP in order to focus cleanup activities in the appropriate area.
- Has a regular street sweeping schedule.[5]

[5] Street sweeping affects trash loading and therefore the sites used for this study will have comparable street sweeping frequencies, to control this variable.

4.0 Data Collection Methods

- On-land visual assessments will be conducted immediately before and after the cleanup activity. Cleanup activities must result in a post-cleanup score of "A" (i.e., low trash generation). Trainings, as well as field checks by the project Quality Assurance Officer, will be mandatory for cleanup crews to learn and maintain the required level of cleanup protocol implementation. Post-cleanup visual assessments will be used to verify that the cleanup protocols are being implemented properly at each site.
- Trash Quantification:
  - Trash will be removed from full trash capture devices and the volume of trash removed will be quantified offsite. Trash will be removed from the DI(s) within an acceptable timeframe before or after the cleanup activity.
  - Trash removed during the cleanup activities will also be quantified offsite.

5.0 Load Reductions

5.1 Calculations

The load reduction for each site will be calculated using two methods, which will be compared for robustness.

Method 1 accounts for the amount of trash removed during each cleanup activity and percent reduction is expressed as follows:

$$\%\ Reduction = 1 - \left(\frac{Vol_{DI}}{Vol_{On-Land} + Vol_{DI}}\right)\ x\ 100$$

Where:
Vol_{DI} = Volume of trash removed from the DI full trash capture device
$Vol_{On-Land}$ = Volume of trash removed by the on-land cleanup activity

Method 2 uses the weighting ratios for different trash generation categories in the reissued Municipal Regional Stormwater Permit (MRP) and provides a higher level of precision than the trash generation rate ranges can predict. An example calculation is provided for Site A, which is a high trash generating rate area with weekly on-land cleanups. The volume of trash measured in the DI is assigned a trash generation rate category and weighting factor per the MRP:

Trash Generation Rate Category	**Trash Generation Rate (gal/acre/year)**	**Weighting Factor**
Low	0-5	0
Moderate	5-10	1
High	10-50	4
Very High	> 50	12

The example calculation assumes five monitoring events have been performed and the results are averaged.

<u>Original Condition: Site A</u>
Baseline Trash Generation Category: High
Weighing Factor: 4
Cleanup Frequency: Weekly

Monitoring Event No.	Trash Generation Rate as Quantified via Trash From DI (gal/acre/yr)	Associated Trash Generation Category After Cleanup	Weighting Factor
1	1.5	Low	0
2	8.6	Moderate	1
3	25.5	High	4
4	7.4	Moderate	1
5	32.5	High	4
	Average Weighing Factor (WF) as after On-Land Cleanup		10/5 = 2

$$\text{Load Reduction} = 1 - \left(\frac{WF_{On-Land\ Cleanup}}{WF_{Baseline}}\right) * 100\% = 1 - \left(\frac{2}{4}\right) * 100\% = 50\%$$

Where:

$WF_{Baseline}$ = Weighting Factor originally assigned to the site

$WF_{On-land\ Cleanup}$ = Average Weighting Factor for trash volume in DI after cleanup events

5.1 Application

It is envisioned that Permittees will use the results of the study to demonstrate trash load reductions from areas where on-land cleanups are implemented consistent with the performance standard that will be developed via the study. Results (i.e., percent reductions) will be applied to jurisdictional land areas and associated baseline trash generation rates that are illustrated on Permittee maps and included in the MRP.

For example, if the study results demonstrate that on-land cleanups performed at a weekly frequency in high trash generation areas effectively reduce trash entering the MS4 by 50%, then an efficiency factor (E_f) of 0.5 is assigned to this control measure implemented at a weekly frequency in a high trash generating area. As described in the MRP, high trash areas have a weighting factor of 4. Applying this efficiency factor to the high trash generating area with a weighting factor of 4 would result in a new/current weighting factor of 2 for this area (i.e., 0.5 x 4), or in other words a 50% reduction of trash from this area. An example trash reduction calculation that applies the scenario described above to a 10-acre area is provided below.

Area where Weekly On-land Cleanups Occur: 10 acres
Baseline Trash Generation Category: High (weighting factor = 4)
Efficiency Factor for Weekly On-land Cleanup Frequency in High Generation Area: 0.5

$$\%\, Trash\ Reduction = \left(\frac{AA\ x\ WF_{Baseline}\ \ x\ E_f}{AA\ x\ WF_{Baseline}}\right) * 100\%$$

Where:

AA	= Area (acres) where Performance Standard is applied
$WF_{Baseline}$	= Weighting Factor originally assigned to the site
E_f	= Efficiency factor for On-land Cleanup Frequency and Trash Generation Category Combination

Attachment #1

Presentation on Results of Preliminary Analysis

On-land Visual Trash Assessments

Data Needed to Establish Baseline Trash Generation and Detect Progress Overtime with Confidence

On-land Visual Trash Assessments

Data Needed to Establish Baseline Trash Generation and Detect Progress Overtime with Confidence

Results of Preliminary Analysis

May 12, 2016

Guiding Questions

- **Baseline Trash Generation**
  - What is the frequency of assessment needed to confidentially establish a baseline trash generation level at an assessment site?
- **Reductions in Trash Generation**
  - What is the frequency of assessment needed to confidentially detect reductions in the levels of trash generated at an assessment site?
- Both depend highly on the temporal variability at assessment sites

Background

- **On-land Assessment Sites (SCVURPPP & SMCWPPP)**
  - Curb, street and sidewalk
  - Randomly selected to avoid biasing
  - Average 1000 feet in linear curb length
  - Associated with:
    - Moderate, high and very high trash generating areas
    - Areas NOT treated by full capture systems

Background

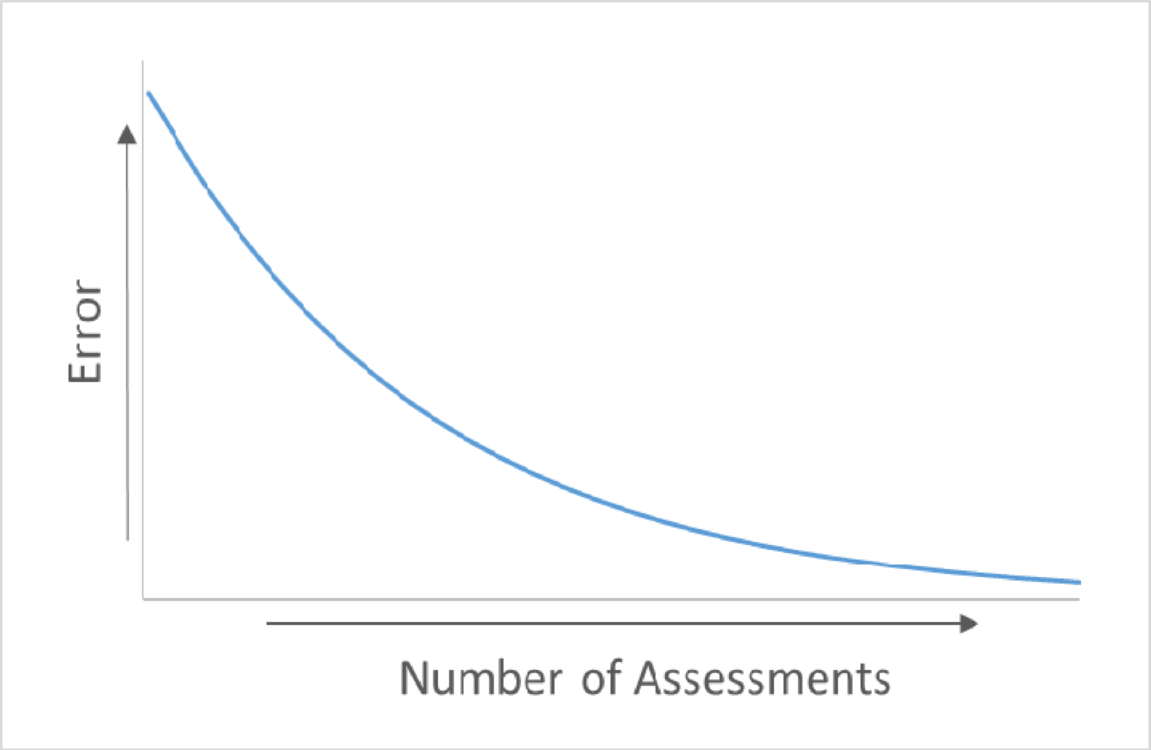
- **Representativeness**
  - **Spatial**
    - MRP 2.0 Goal = At least 10% of linear curb feet (street miles) assessed in TMA
    - Key is that baseline trash generation in linear curb feet assessed are generally (+/- 2%) proportional to the trash generation in TMA
    - Example – Proportion of acres and linear feet assessed:

Baseline	Low	Mod	High	Very High
Acres	NA	60%	30%	10%
Linear Curb Feet Assessed	NA	58%	31%	11%

Background

- **Representativeness**
  - **Temporal**
    - Baseline Trash Generation
    - Observing progress/change over time

Error in Trash Assessment Score



Statistical Analysis – Establishing Baseline

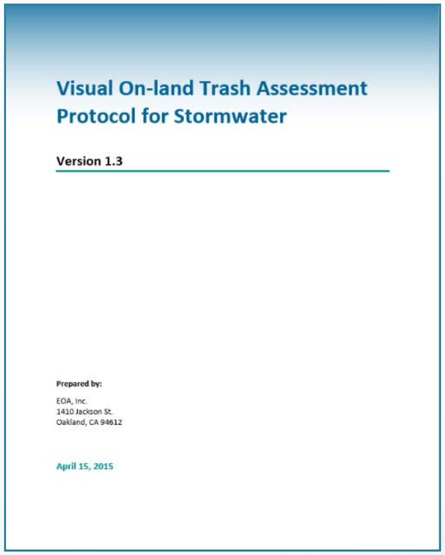
- **Establishing Baseline** - What is the frequency of assessment needed to confidentially establish a baseline trash generation level at an assessment site?
- **Analysis Methods**
  - Subsample most frequently assessed sites to determine level of error in baseline generation designation:

Error = Actual Score – Observed Score

- Where:
  - ***Actual Score*** = *average assessment score determined by maximum # of assessments*
  - ***Observed Score*** = *average assessment score determined by a smaller (subset) # of assessments*

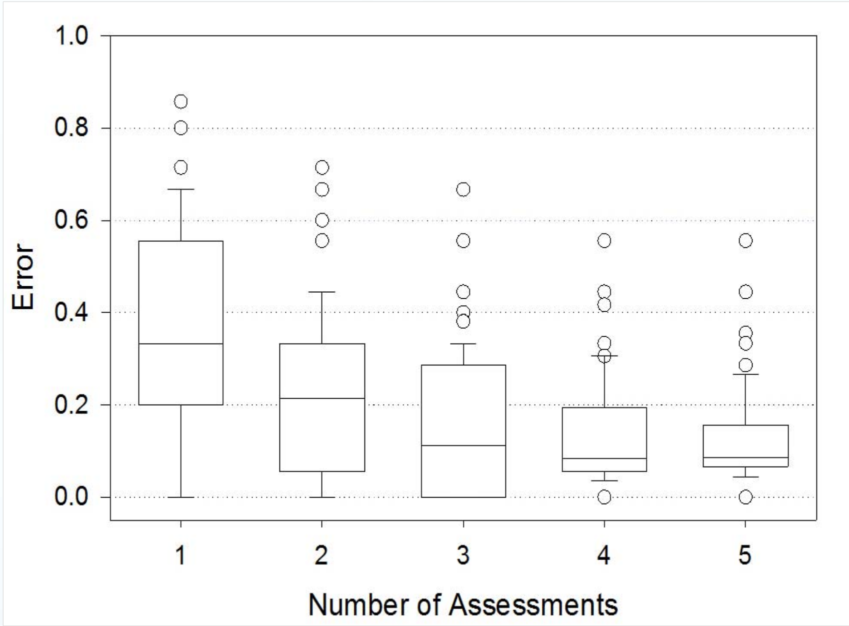
Assessment Dataset Utilized - Baseline

- **Tracking California's Trash Project**
  - 6-9 months of data at each sites
  - Only assessments that occurred mid-way between street sweeping were used
  - Wet and dry seasons
  - Assessments Sites and Frequencies
    - Fremont – 2 sites, 7 assessments each
    - Oakland – 4 sites, 9 assessments each
    - San Jose – 6 sites
      - 2 with 9 assessments
      - 2 with 7 assessments
      - 2 with 5 assessments



Baseline Analysis Results

- Assessment scores transformed into numerical values
  - 0 = low
  - 1 = moderate
  - 2 = high
  - 3 = very high
- Error
  - 1.0 = Full assessment score
    0.5 = ½ assessment score
- 1 assessment = 90% chance +/- 0.67 error
- 2 assessments = 90% chance +/- 0.44 error
- 3 assessments = 90% chance +/- 0.33 error



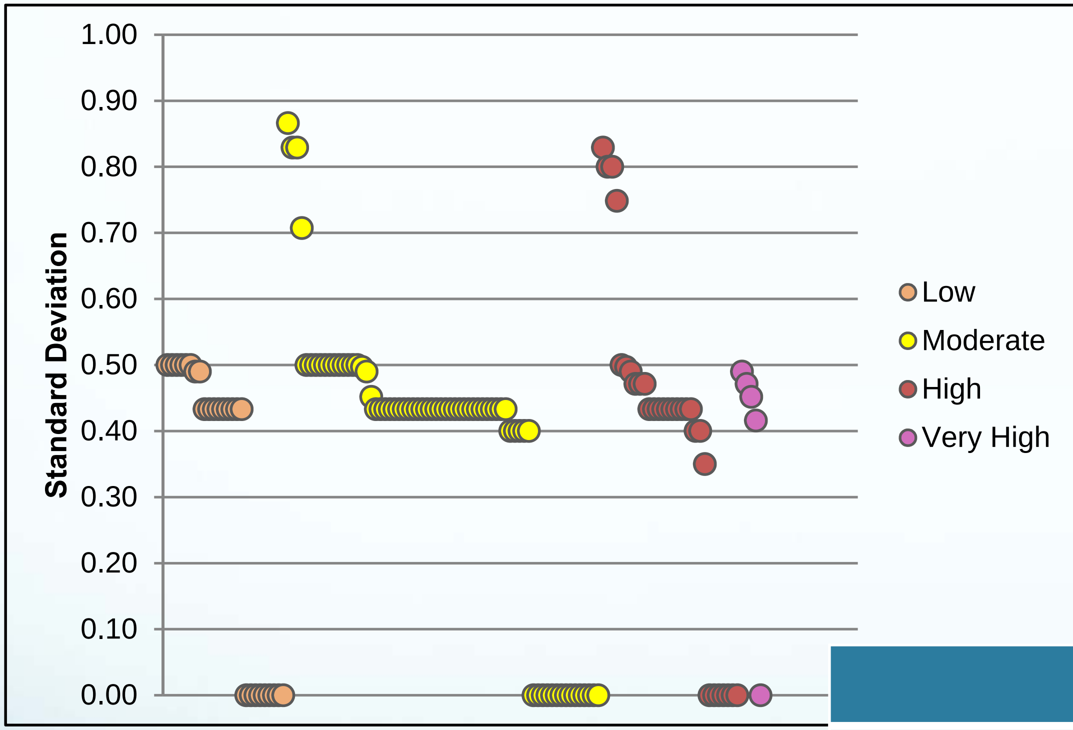
Statistical Analysis – Detecting Reductions in Generation

- **Observing Change** - What is the frequency of assessment needed to confidentially detect reductions in the levels of trash generated at an assessment site?
- **Analysis Method**
  - Power analysis to determine adequate # of assessments need to have 90% confidence that an assessment score (i.e., trash generation) has changed:

$$n = \frac{(\sigma_1^2+\sigma_2^2)\left(z_{1-\frac{\alpha}{2}}+z_{1-\beta}\right)^2}{\Delta^2}$$

Where:
n = sample size for each group (assuming the groups have the same sample size)
σ_1, σ_2 = standard deviation of dataset #1 and #2
$\Delta = |\mu_2-\mu_1|$ = the absolute difference between two means
α = probability of type I error
β = probability of type II error
z = Z score for a given α or β

Differences in Temporal Variability between Generation Categories?

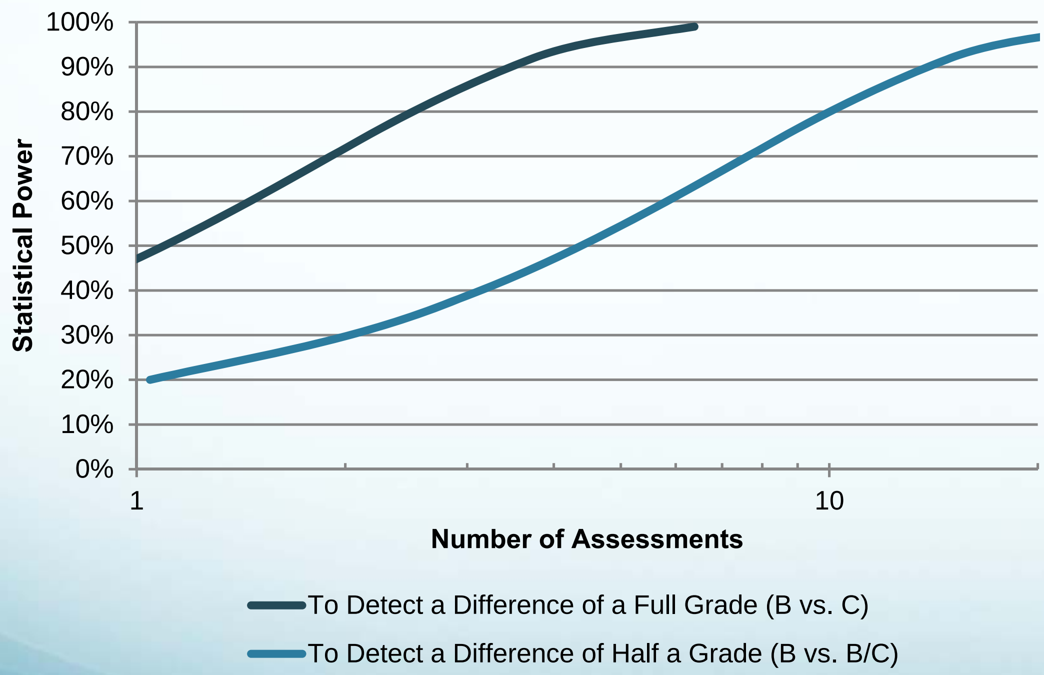


- Comparison of standard deviations (variability) of assessment scores
- Grouped by trash generation category
- Sites with 4 or more assessments (n = 129)
- No significant differences in temporal between generation categories

	Low	Mod	High	Very High	All Sites
Count	26	68	29	5	129
Max	0.50	0.87	0.83	0.49	0.87
Median	0.43	0.43	0.43	0.45	0.43
Mean	0.30	0.37	0.40	0.37	0.36
Min	0.00	0.00	0.00	0.00	0.00

Assessment Datasets Utilized & Results
Detecting Change

- **SCVURPPP and SMCWPPP Assessments**
  - Standard deviation of 129 sites assessed ≥ 4 times (0.45)
  - 90% statistical confidence



Power (likelihood of detecting a difference when one exists)	**Number of Assessments Required to Meet a Given Statistical Power (σ = 0.45, 90% confidence)**	
	To Detect a Difference of a Full Grade/Score (e.g., C to B)	To Detect a Difference of ½ a Grade/Score (e.g., C to B/C)
80%	3	10
90%	4	14

Summary

- **Spatial Representativeness**
  - Baseline trash generation in the linear curb feet assessed should be proportional to acreage assessment results are applied to
  - Following this approach, assessments of 10% of curb feet should suffice for most TMAs, with the exception of very small TMAs (i.e.,< 10,000 linear feet)

Summary

- **Temporal Representativeness**
  - Establishing Baseline
    - On average, a 90% chance of being within 1/2-2/3 of an actual generation category by conducting 1 or 2 assessments ½ way between street sweeping or other reoccurring (existing) controls
    - Accuracy increases when adding additional assessments
  - Detecting Change/Progress
    - Variability similar among trash generation categories; different assessment frequencies not needed
    - Adequate assessment frequencies
      - To have a 80% likelihood (with 90% confidence) of detecting:
        - Change in 1 trash generation category – 3 assessments
        - Change in ½ trash generation category - 10 assessments