Presentation to SWRCB
July 28, 2006

Gary Lorden, Ph.D.

Statistical Consultant

Panel's "Statistical" Comments

- "…there is wide variation in stormwater quality from place to place, facility to facility, and storm to storm (p.6)."

- "Since the storm-to-storm variation at any outfall can be high, it may be unreasonable to expect all events to be below a numeric value." (p.6)."

- "…several to more times each year, the runoff volume or flow rate from a storm will exceed the design volume or rate capacity of the BMP. Stormwater agencies should not be held accountable for pollutant removal from storms beyond the size for which a BMP is designed." (p.10)

- "The Panel recommends that Numeric Limits and Action Levels not apply to storms of unusual event size and/or pattern (e.g. flood events)." (p.18)

What kinds of statistics are needed?

- **Design of data collection needed to choose and maintain numerical limits— range of storm variables, receiving water variables, flows/volumes/dilutions, sufficient sample sizes and locations**

- **Statistical profile of a facility's data distributions of effluent constituents, in relation to storm variables, etc. Is "lognormal distribution" model sufficient? Indications are negative.**

- **Useful estimates of "frequency of exceedance" for possible choices of numerical limits**

- **Statistical procedures for updating estimates– and possibly modifying numerical limits– as data accumulate over a period of years (with year-to-year variation in number and severity of storms).**

- **Establish data-taking requirements for monitoring and compliance: How much data is needed?**

Sources of Data Variability

- **Input Variables**
  -- Influent characteristics
  -- Storm characteristics (e.g. rainfall intensity, flow rate/volume)
  -- Site-specific hydrologic features
  -- Receiving water characteristics (e.g. $\rightarrow$ dilution)

- **Treatment Characteristics**
  -- Treatment capabilities (facilities + inputs $\rightarrow$ great variability in quality of treated effluent)

  -- Flow rate/volume-based capacity limits

- **Output Variables**
  -- Measurements of discharge constituents
  -- One grab-sample/storm vs. hourly averages vs. storm composite concentrations
  -- Lab analysis screening strategy

Sample Dataset: Grab sample measurements of copper at three outfalls

(water quality objective = 14 mg/L)

	A	B	C
Sample Size	23	32	20
Sample Median	2.8	2.8	3.2
Sample Maximum	55	12	39
Probability*	.003	.012	.110

- Probability of obtaining a maximum "as large as this" if data distribution was lognormal
- The largest values in such datasets are typically "too large to fit a lognormal model". Therefore, statistical analysis based on a lognormal model is not justified.

How large do sample sizes need to be?

- Failure of lognormal model
  - ➔ Cannot use calculated "average" and "standard deviation" of a dataset to estimate the frequency of "over-the-limit" events (or "exceedances") for a given numerical limit– since the "shape" of the distribution is unknown

- Fallback: use the "observed exceedances" in the dataset
  - ➔ very large sample sizes are needed

- Example: To show that the exceedance rate is at most 5% requires, for example, a sample size of 153 with at most 3 exceedances. (For "95% confidence".)

- "Caesar's wife" effect: to reliably avoid the *suspicion* of "having more than 5% exceedances", one must have *much less than* 5%-- e.g. a 1.1 % *true exceedance rate* is needed to guarantee a 90% probability of *avoiding suspicion* when the sample size is 153.

Key Statistical Points

- **Important to base numeric limits on data collection designed to reflect the full ranges of major sources of variability in effluent measurements -- storms, receiving waters, facilities/treatments, and sampling methods**

- **Need updating of data to reflect year-to-year variability of storms– and corresponding reassessment of numerical limits**

- **Don't rely entirely on highly variable data like grab samples or optimistic statistical assumptions like "lognormal"**

- **Data requirements for compliance monitoring are severe— should use large enough sample sizes and shouldn't impose "too high expectations of Caesar's wife"**