Date: August 16, 2019

From: John Fox, US Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Washington Division
fox.john@epa.gov Alternate work location phone (703) 838-1677

To: Rich Breuer, Environmental Program Manager, Division of Water Quality, California State Water Resources Control Board (Rich.Breuer@Waterboards.ca.gov)

CC: Robyn Stuber, USEPA-R9 Water Division
Debra Denton, USEPA-R9 Water Division

Subject: Probability of Failing TST and WET Maximum Monthly Effluent Limit

This is an update to the draft memo which I sent you on May 21, 2019, in response to your email of May 8, 2019. There have been some minor corrections, and some numbers have changed slightly because simulations were redone. This memo examines the probability of various events that might occur during routine WET compliance monitoring under the provisions recently proposed by the California State Water Resources Control Board (2019). Please contact me if anything requires further explanation.

I examined these probabilities by assuming various "false-positive" rates for the Test of Significant Toxicity (TST). Higher false-positive rates are associated with greater within-test variability and insufficient replication in WET tests. These error rates are described in a recent publication (Fox et al. 2019). They are based on probability calculations using assumed values for the percent effect (PE) at the IWC, and various coefficients of variation (CV) for within-test variability.

The relevant events appear in the Draft Provisions on page 23 (*iv. MMEL Compliance Tests*), page 27 (*(B) Chronic Aquatic Toxicity MMEL*) and page 28 (*f. Toxicity Reduction Evaluation*). This memo addresses only sublethal endpoints of WET tests for chronic toxicity; it does not consider survival outcomes in either chronic or acute WET tests. This memo evaluates only what may happen when toxicity is low (true percent effect 10% or less). MDEL violations are therefore not considered, because their probability is very low when toxicity is low (Appendix).

Outline of events leading to a TRE

I assume that routine monthly WET compliance monitoring occurs for 5 years (60 months).

- TST-test "Pass" at the instream waste concentration (IWC) during routine compliance monitoring: routine testing continues next month
- TST-test "Fail" at the IWC during routine compliance monitoring: conduct another test ("MMEL Compliance Test #1") on a new sample in the same month
    - MMEL Compliance Test #1 is a "Fail": record an MMEL violation
    - MMEL Compliance Test #1 is a "Pass": conduct another test on a new sample

- MMEL Compliance Test #2 is a “Fail”: record an MMEL violation
- MMEL Compliance Test #2 is a “Pass”: No violation

In words: if a TST “Fail” occurs during routine compliance monitoring, it triggers at least one and possibly two additional tests on new samples in the same month, and if a “Fail” occurs for either of these, an MMEL Violation occurs.

If MMEL violations occur in two successive months [called a run of length 2, below], a TRE is triggered [page 23, *f. Toxicity Reduction Evaluation*].

Outline of Required Calculations for Non-Toxic Samples

To calculate the risk of needing to conduct a TRE, the following probabilities are required:

- probability of a TST “Fail” occuring for a single WET test
- probability of a sequence of events leading to a MMEL violation during a single “month”
- probability of 2 successive violations [run of 2 violations] occuring at least once in 5 years

These probabilities will be determined for a hypothetical case corresponding to negligible effects on survival and the sublethal endpoint. In particular, a negligible effect (acceptable effect) on the sublethal endpoint is defined as a 10% effect, i.e., the IWC mean is 90% of the control mean. The 10% reduction for a sublethal endpoint is based upon USEPA (2010), where a ten percent effect on a sublethal endpoint is deemed to be ‘negligible.’ A false positive occurs when a sample is incorrectly declared toxic when the true effect level is at or below a value deemed to be “acceptable” (i.e., true percent effect at the IWC is $\leq$10%).

In addition to determining the probability that 2 successive violations occuring at least once in 5 years, we also want to determine (a) the expected number of violations that might occur in 5 years (or the distribution of this number) and (b) the expected number of runs of 2 violations in 5 years (and, if possible, the distribution of the number of such runs).

Calculations herein will rely on the control coefficient of variation (CV) for the sublethal endpoint for *Ceriodaphnia*.

Calculations were made using R (R Core Team 2017). The R functions are documented in the text file “R-functions-MMEL.R”, which is a plain ASCII text file and which is transmitted with this memo.

Summary of Results

The probability of a TST false positive during routine compliance monitoring depends on the within-test variability (here quantified by the control CV) and number of replicates used in the WET test. For the chronic *Ceriodaphnia* test and its sublethal endpoint of reproduction, the median control CV is around 0.15, as determined from a large number of tests and laboratories (Table 2 of Fox et al. 2019).

For a control CV of 0.15, the probability of a TST false positive is about 0.05 when using the minimum of 10 replicates (Table 1a, below), 0.003 for 20 replicates (Table 1b), and <0.001 for 30 replicates (Table 1c). A few laboratories may have a median control CV as high as 0.21 to 0.24 (Table 3 of Fox et al. 2019) which would necessitate either decreasing within-test variability and/or increasing replicates to achieve a false-positive probability under 0.10 (Table 1). It follows that not all laboratories will achieve a false-positive rate ≤0.05 (unless they increase the number of replicates) but most are expected to achieve TST false-positive rates less than 0.10 using their current number of replicates. Therefore, in this analysis we desired to be cautious and chose a false-positive probability of 0.10 rather than 0.05 as conservative point of reference for evaluating the risk of an MMEL violation.

This analysis shows that if the TST false positive probability is ≤ 0.10, the probability of a MMEL violation in any one month is less than 0.02 (i.e., less than 2 percent of the time) and the probability of two successive MMEL violations in 60 months is no more than 0.02 (Table 2). If the TST false positive probability is ≤ 0.05, the probability of a MMEL violation in any one month is less than 0.005 (i.e., less than ½ of one percent of the time) and the probability of two successive MMEL violations in 60 months is no more than 0.0011 (Table 2).

Probabilities of failing TST

A false positive occurs when a WET-test sample is incorrectly declared toxic ("Fail" for TST) when the true effect level is at or below a value deemed to be "acceptable," that is, when the true percent effect ("PE") at the IWC is ≤10%.

Table 1 shows the probability of declaring a WET test sample to be toxic (i.e., showing a statistically significant decrease in Ceriodaphnia reproduction) using the TST approach. These data can be used to calculate the probability of having 2 successive MMEL violations. The values of PE and CV represent parameter values.[1]

These are exact calculations of statistical power (not simulations). The calculations are based on a set of assumptions (e.g., the standard deviation of the sublethal endpoint was the same for IWC and Control) and should not be considered exact for actual WET tests.

Data in Table 1 were produced by R function *TST.pwr.fn2*, posted originally at https://figshare.com/articles/WET_Error_Rates_for_TST_NOEC_Supporting_Information/7122812 as a supplement to Fox et al. (2019). The function is included in the supporting document "R-functions-MMEL.R", transmitted with this memo.

1 This PE value refers to the parametric value, not a sample estimate's value. Sample estimates can differ markedly from parameter values; accurate estimates require 30-100 data points. The parametric value for PE is $(1 - \mu_{IWC} / \mu_{CON})$ where μ_{IWC} and μ_{CON} are the mean parameters for IWC and Control, respectively. The parametric values for CV are $(\sigma_{IWC} / \mu_{IWC})$ and $(\sigma_{IWC} / \mu_{CON})$.

Table 1a. Probability of failing TST for specified parameters *PE* and *CV*, using 10 replicates				
PE:	0%	10%	25%	50%
CV:				
0.10	0.0000	0.0020	0.8000	1.0000
0.15	0.0000	0.0480	0.8000	1.0000
0.20	0.0110	0.1500	0.8000	1.0000
0.30	0.1070	0.3410	0.8000	0.9980
0.40	0.2350	0.4610	0.8000	0.9920

Table 1b. Probability of failing TST for specified parameters *PE* and *CV*, using 20 replicates				
PE:	0%	10%	25%	50%
CV:				
0.10	0.0000	0.0000	0.8000	1.0000
0.15	0.0000	0.0030	0.8000	1.0000
0.20	0.0000	0.0340	0.8000	1.0000
0.30	0.0170	0.1740	0.8000	1.0000
0.40	0.0830	0.3110	0.8000	0.9990

Table 1c. Probability of failing TST for specified parameters *PE* and *CV*, using 30 replicates				
PE:	0%	10%	25%	50%
CV:				
0.10	0.0000	0.0000	0.8000	1.0000
0.15	0.0000	0.0000	0.8000	1.0000
0.20	0.0000	0.0070	0.8000	1.0000
0.30	0.0030	0.0900	0.8000	1.0000
0.40	0.0300	0.2130	0.8000	1.0000

Probability of a Violation of MMEL in any one Month

Notation:

p_i = Prob{Fail TST | CV, PE, nreps}, i = 1,2,3

i = 1,2,3 denote successive WET tests, the first being a routine monthly compliance monitoring test, the 2nd and 3rd being MMEL compliance tests

Notice that CV, PE, and nreps can differ for each test; the notation (p_i) is a reminder of this. Numerical examples below, and simulations, assume they are constant (i.e. that p_i all equal one value, p).

Assuming independence of successive TST failures with probabilities p_1, p_2, p_3:

P_V = Prob{one MMEL violation in any month} = $p_1 * p_2 + p * (1-p_2) * p_3$

The events represented as occuring in one month in the preceding equation are:

Fail TST in compliance monitoring (p_1)

Followed by failure of the first MMEL compliance test: $p_1 * p_2$

if the 1st MMEL compliance test is passed, the 2nd may be failed: $p_1 * (1-p_2) * p_3$

For subsequent calculations we will assume that $p = p_2 = p_3 = p$ and that p represents the probability of a false positive for TST.

Numerical examples:		P_V
$p = p_2 = p_3$ =0.05 :	0.05*0.05 + 0.05*0.95*0.05 =	0.004875
$p = p_2 = p_3$ =0.10 :	0.10*0.10 + 0.10*0.90*0.10 =	0.019000
$p = p_2 = p_3$ =0.15 :	0.15*0.15 + 0.15*0.85*0.15 =	0.041625
$p = p_2 = p_3$ =0.20 :	0.20*0.20 + 0.20*0.80*0.20 =	0.072000
$p = p_2 = p_3$ =0.25 :	0.25*0.25 + 0.25*0.75*0.25 =	0.109375

Probability of 2 Successive MMEL Violations in 60 Months (Triggering a TRE)

The probability of back-to-back MMEL violations in two successive months is calculated using (a) the probability of a MMEL violation in any one month, P_V, above, which is based on the probability of failing TST in routine monthly compliance monitoring. Occurrences of MMEL violations were assumed to be independent from month to month. Calculations and simulations were done in the R statistical computing language (R Core Team, 2017). Scripts for the functions are provided separately.

Obtaining the probability of MMEL violations in adjacent months required simulation. When this probability is small, the expected number of runs of length ≥ 2 is equivalent to this probability (because then such runs are infrequent, so only one or no such runs occur in 60 months and the run, if it occurs, is almost always of length 2). However, as this probability increases, the expected number of runs will differ from the desired probability, so the probability was calculated by simulation (while simultaneously verifying that the simulations agreed with the following equation for number of runs of length ≥ 2). The simulation logic and calculations were also verified in detail using numerical examples.

The expected number of runs of length ≥ 2 for heads (successes) in a sequence of *n* Bernoulli trials with heads (success) probability *p* is (Schilling 1990, Bloom 1996):

$n_R = n*(1-p)*p^2$

In this equation, *p* is identified with P_V (the probability of MMEL failure in any one month).

Table 2. Probability of at least one run of 2 or more MMEL failures

Probability *p* of failing TST during routine monitoring	P_V[1], probability of MMEL violation each month	Probability of one or more runs (2 successive MMEL failures) in 60 months, resulting in TRE (based on P_V)[2]
0.05	0.004875	0.0011
0.10	0.019000	0.020
0.12	0.027072	0.042
0.13	0.031603	**0.056**
0.15	0.041625	0.093
0.20	0.072000	0.25
0.25	0.109375	0.48

[1] $P_V = p*p + p*(1-p)*p$

[2] Average of six simulation runs, each consisting of 10,000 simulated sequences of 60 months

A run of MMEL failures leading to a TRE becomes more likely as the TST false positive rate increases (Table 2). If a TST false positive is defined by failing TST when the true percent effect is 10% (or less), then it is desirable to keep this probability under 0.13 in order to keep the risk of 2 successive MMEL compliance failures less than 0.05 over the course of 60 months (notice figures in bold). This may necessitate either decreasing within-test variability and/or increasing replicates. depending on the laboratory's within-test CV (Table 1).

Calculations assumed no serial correlation among tests (routine monthly compliance tests and MMEL compliance tests). Serial correlation might be expected if toxicity appeared in one month and continued into the next month, but that is inconsistent with the false-positive scenario considered here. Serial correlation could conceivably occur without toxicity, if caused by (for example) a sequence of tests with high CVs or impaired test organisms. However, such anomalous conditions can readily be detected by examination of test data and comparison to recent test data and control charts.

References

Bloom, DM (1996) Probabilities of Clumps in a Binary Sequence. Mathematics Magazine 66(5):366-372, downloaded on 2019-02-16 from: https://www.maa.org/programs/faculty-and-departments/classroom-capsules-and-notes/probabilities-of-clumps-in-a-binary-sequence

California State Water Resources Control Board (2019) First Revised Draft, Water Quality Control Plan For Inland Surface Waters, Enclosed Bays, and Estuaries of California, July 25, 2019. 49 pages. Accessed on 2019-07-30 at https://www.waterboards.ca.gov/water_issues/programs/state_implementation_policy/docs/toxicity_2019_provisions_1strevdraft.pdf

Fox, JF, DL Denton, J Diamond and R Stuber. (2019). Comparison of false positive rates of two hypothesis-test approaches in relation to laboratory toxicity test performance. Environmental Toxicology and Chemistry 38(3):511-523. https://doi.org/10.1002/etc.4347

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Schilling, MF (1990) The Longest Run of Heads. The College Mathematics Journal 21(3):196-207, downloaded on 2019-02-16 from: https://www.maa.org/programs/maa-awards/writing-awards/the-longest-run-of-heads

USEPA (2001). Final Report: Interlaboratory Variability Study of EPA Short-term Chronic and Acute Whole Effluent Toxicity Test Methods, Vol. 1. United States Environmental Protection Agency, Office of Water. EPA 821-B-01-004, September 2001.

USEPA (2010). National pollutant discharge elimination system test of significant toxicity technical document. EPA/833/R-10/004. Washington, DC.

APPENDIX

I. MDEL violations and the probability of observing survival of 50% or less in non-toxic samples

An MDEL violation occurs when the sublethal endpoint 'fails' the TST and the observed percent effect for the survival endpoint is ≥ 50%. Note that percent effect for survival ≥50% implies survival <50%.

Available evidence suggests that the probability of observing survival of <50% in non-toxic samples is negligible. The EPA Inter-Laboratory Variability Study (USEPA 2001) provides relevant evidence (Table 3).

The LC50 for survival was always ≥100% for blank samples and for all but one reference toxicant test. The reference toxicant experiments produced IC25s ">100%" for 72% of tests with an average IC25 of 86%. Thus, the reference toxicant samples had low sublethal toxicity to *Ceriodphnia* and also did not produce an LC50 (and therefore an estimated <50% survival) within the range of tested toxicant concentrations, with one exception.

Also, control survival was uniformly high (last three columns). This is informative because it can serve as a surrogate for a nontoxic IWC concentration. That is, it provides evidence that random occurrences of low survival in nontoxic samples are rare. Survival would need to be <50% in the IWC to result in a percent effect for survival ≥50%. Considering both valid and invalid tests (some of which were invalid because control survival was <80%), for only 2/109 tests was control survival <50% (the values were 30% and 50%)

Table 3. Summary of results from USEPA (2001), Tables 9.7, 9.8, and 9.9 for the *Ceriodaphnia* chronic test method.

Sample Type	Number of Valid (Invalid) Tests	Number of tests with survival LC50 ">100%" in test sample:	Number of valid tests plus invalid tests with survival in control:		
			100%	<80% and ≥50%	> 50%
Blank	27 (7)	27	30	4	0
Reference Toxicant	37 (11)	36 [a]	44	3	1
Effluent Samples	24 (3)	0	27	0	0

[a] One test had LC50 at 55.8%

Probability Analysis for MMEL Limit

Description of MMEL Limit

The Median Monthly Effluent Limit (MMEL) depends only on TST outcomes for the sublethal endpoint of the required WET test method; it does not use the percent effect. TST results can be coded 0 (pass, when TST null hypothesis is rejected) and 1 (fail, when the TST null hypothesis is not rejected). The MMEL is exceeded if the median of coded TST pass-fail results for the month is positive. The MMEL median is based on not more than 3 WET tests in a calendar month (one scheduled test and possibly one or two followup tests).

Terminology:

- Scheduled WET test ("routine monitoring"; may be monthly or less frequent)
- Followup WET test (occurs only when scheduled test fails TST).
- Individual WET tests are classified Pass (0) or Fail (1) on the basis of the TST result
- The MMEL is said to be 'exceeded' or 'not exceeded'

A mutually exclusive and exhaustive list of possible events for MMEL determination follows:

- First WET test (scheduled WET test):
    - "pass TST", stop, MMEL is not exceeded. Next test is a scheduled WET test.
    - "fail TST", conduct second WET test immediately (followup test).
- Second WET test (followup test):
    - "fail TST", stop, MMEL is exceeded (the median cannot be made zero by conducting a third WET test).
    - "pass TST", conduct third WET test immediately (followup test).
- Third WET test (followup test):
    - "pass TST", MMEL is not exceeded.
    - "fail TST", MMEL is exceeded.

Caveats

Results calculated herein can be interpreted as approximations to the probabilities of the events described. Bear in mind that these proportions are not exact and that achieved probabilities of these events may differ from calculated estimates of probabilities when (a) probabilities are small, (b) replicate data are not well-approximated by the normal distribution, (c) other assumptions do not match behavior of real data. These remarks are made in order to caution against assuming that the results reported here apply *exactly* to any given situation, based on values estimated from data.

In particular, CV values estimated from WET-test data will differ from the true, parametric value. Getting an accurate estimate of the true CV requires data from many WET tests, and such estimates should always be reported with a confidence interval to indicate a probable range for the true value (this is not strictly correct; a Bayes posterior credible interval would better achieve this).

The long-run average CV cannot be estimated with high precision from samples of, say, 10-20 WET tests, and its true value might change over time. Also, the CV for the tested water sample needs to be considered. A cautious (perhaps overly cautious) approach would be to calculate an upper percentile (like the 75th or 80th percentile) of 10-20 values for the CVs of control and RWC, using the larger of the percentiles.

All analyses below address the probabilities of failing TST and exceeding the MMEL in a sequence of scheduled WET tests conducted using the *Ceriodaphnia dubia* survival and reproduction test method, in particular using the reproduction endpoint (survival *per se* is not considered).

All analyses below assume that the probability of failing the TST (i.e., not rejecting the TST null hypothesis) is constant. Variation in this probability, such as the occurrence of transient toxicity, is not considered here.

All analyses below assume that all conducted WET tests are valid and that the number of replicates is constant (either 10, 20, or 30, depending on the context).

False Positives and True Positives for TST and MMEL

False positives for MMEL are defined as MMEL exceedances that occur as a result of false positives for TST.

TST false positives can be defined as TST 'fails' that occur when the true (parametric) percent effect is not more than 10%. A true percent effect of at least 25% can be used as a working threshold for a true positive. Toxicity occurs if the true percent effect is between 10% and 25%, but for practical reasons, as a regulatory management decision, it was useful to select a value of 25% percent effect to use in the TST t-test (USEPA 2010), and that convention is used here. The TST false positive rate depends on within-test variability (USEPA 2010; Fox et al. 2019). Thus, one must consider both percent effect and variability (quantified here by the coefficient of variation for within-test variability of the sublethal endpoint).

Tables of expected probabilities for TST fails are provided separately in spreadsheets named "ptst.n102030.sdconst.csv" and "ptst.n102030.cvconst.csv". The first used the same standard deviation (SD) for controls and water samples (RWC). The second used the same coefficient of variation (CV) for the two samples. These two alternative approaches were discussed by Fox et al. (2019) and the choice between them should have little influence on results when percent effect is 10% or less. Calculations are based on power calculations for the t-test, using the non-central t-distribution. These were generated using the R function "TST.pwr.fn2" (R functions are provided in a text file,"R-functions.txt").

Probabilities of False and True Positives for TST and MMEL

Three events seem to be of primary interest: (1) "fail' outcomes for TST, because these result in followup testing, (2)MMEL exceedances, and (3) back-to-back MMEL exceedances, because these can result in a TRE.

These 3 events are related to the parameters that determine probability of failing TST: the true percent effect and the coefficient of variation (CV) of Ceriodaphnia reproduction measurements. As noted above, tables are provided in spreadsheets showing the probability of failing TST in relation to values of these parameters.

Therefore, the probabilities of the other two events, MMEL exceedances and back-to-back MMEL exceedances, will be tabulated for selected values of the probability of failing TST rather than values of CV and percent effect. However, graphics and spreadsheets (provided separately) will be used to relate these two MMEL outcomes to CV and percent effect.

Table 1 shows the probabilities of MMEL exceedances for TST fail probabilities of 0.02 to 0.40. The TST fail probabilities in the first column cover the most common values for a wide range of CVs and percent effect values of interest. Calculations assume that the probability of failing TST is the same for each and every WET test. The third column shows the probability of back-toback MMEL exceedances.

***Table 1. Probabilities of MMEL exceedances**.*

Probability of failing TST	Probability of exceeding MMEL	Probability of back-to-back MMEL exceedances
0.02	0.000792	6.27e-07
0.04	0.00314	9.83e-06
0.06	0.00698	4.88e-05
0.08	0.0123	0.000151
0.1	0.019	0.000361
0.12	0.0271	0.000733
0.14	0.0365	0.00133
0.16	0.0471	0.00222
0.18	0.059	0.00348
0.2	0.072	0.00518
0.22	0.0862	0.00742
0.24	0.101	0.0103
0.26	0.118	0.0138
0.28	0.135	0.0182
0.3	0.153	0.0234
0.32	0.172	0.0296
0.34	0.192	0.0368
0.36	0.213	0.0452
0.38	0.234	0.0547
0.4	0.256	0.0655

Followup Testing

The frequency of TST-fails (during scheduled monitoring) translates to the frequency of followup monitoring. When the probability of failing TST is low, a TST-fail at a scheduled monitoring time will usually involve two followup tests (because the first followup test is likely to be a 'pass'). However, in this case the probability of exceeding the MMEL is quite small (Table 1).

If a TST-fail during scheduled monitoring results from toxicity, the first followup test might have a high probability for TST-fail, assuming that the toxicity persists for a week or more. The analyses reported here do not take account of transient toxicity events. Doing so would require a different set of analyses. These analyses focus on constant, low probabilities of failing the TST.

Probability of MMEL Exceedances Based on Monitoring Frequency

Permits may require monitoring monthly, quarterly, semi-annually, or annually, depending on a determination by the permitting authority. Thus scheduled monitoring can result in a long sequence of WET tests, and some of these may result in exceedances.

The question may arise, 'what is the chance of an MMEL exceedance in any given year if monitoring is conducted quarterly or semiannually as compared to monthly monitoring?'

The probability of at least one MMEL exceedance can be calculated using the binomial distribution (Appendix 3). A similar calculation applies to the probability for back-to-back exceedances.

The more frequent monitoring is, the more likely it is that an exceedance will occur. This applies to false positives and to true positives (though these will occur with very different probabilities).

When true percent effect is large, it is desirable that MMEL exceedances occur with high probability. Conversely, when true percent effect is small, it is desirable that MMEL exceedances occur with low probability, because these would be false-positives.

Figure 1 provides results for some representative values of percent effect and control CV ('cCV') and numbers of replicates ('N'), and for three representative values of monitoring frequency. Figure 1 shows the probability that at least one exceedance of MMEL will occur during a sequence of 2 (dotted line), 4 (dashed), and 12 (solid) scheduled WET tests. The percent effect and CV values are parameters (aka 'true values'), not sample estimates observed in WET tests. Calculations for probability of TST failure assumed that the standard deviation is the same for the distributions of reproduction in control and water (RWC) samples.

Figure 1 does not apply to back-to-back MMEL exceedances (and the risk of starting a TRE). It shows the probability of at least one exceedance. Figure 2 shows the probability of at least one occurrence of back-to-back MMEL exceedances during a sequence of 2, 4, or 12 scheduled WET tests.

Figure 1. Probability that an exceedance of MMEL occurs during scheduled monitoring.

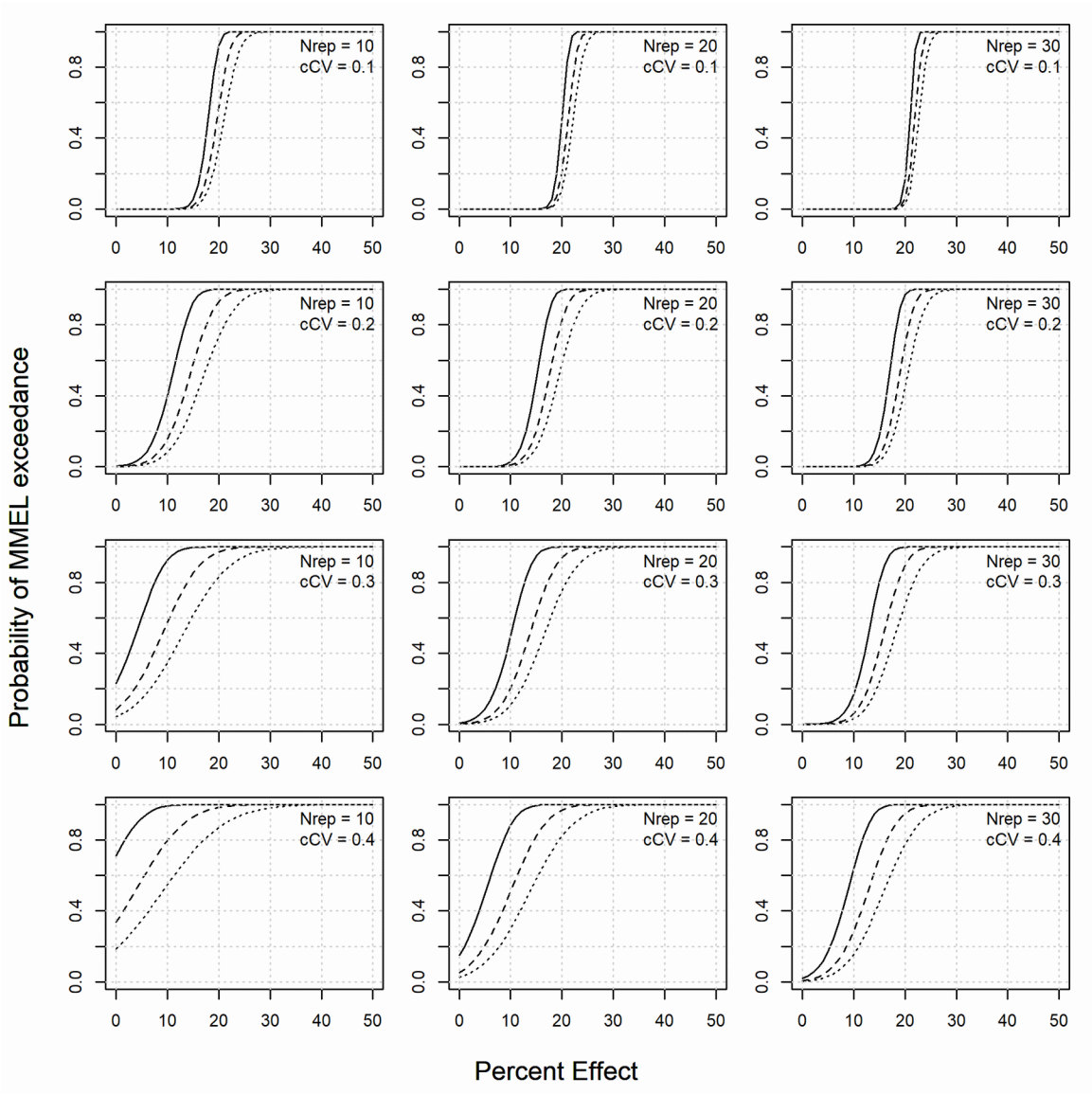
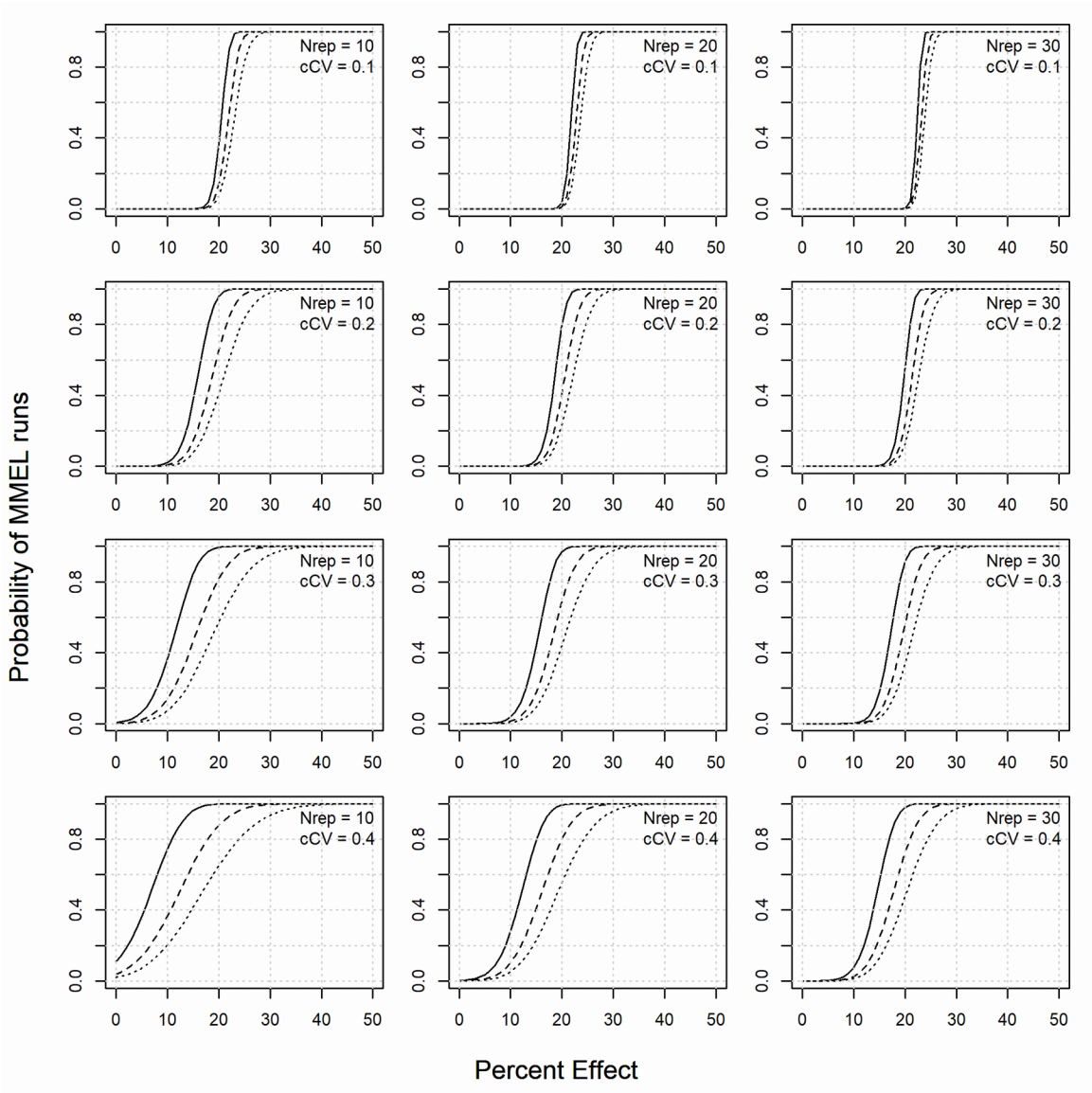


Figure 2. Probability that a back-to-back MMEL exceedance occurs during scheduled monitoring.



Appendix 1: Concepts for Simulating MMEL Exceedances

Concepts are first described in terms of simulated sequences of events, later as probability calculations. This will provide details of the reasoning used as a basis for calculations. It is intended to help the interested reader understand the concepts and calculations in a way that a peremptory statement of the equations given later would not. It should also provide readers with enough information to decide whether the concepts are correct (R functions implementing these calculations are provided separately and are commented to aid review).

Calculations assume outcomes of all WET tests are independent of one another and that the probability of failing TST remains constant.

Simulated Sequences of WET Tests

A sequence of scheduled WET tests and associated followup tests can be illustrated by the schematic below (0 = Pass TST, 1 = Fail TST, x=no followup test needed).

```
Scheduled Tests:    0   0   1   0   0   1   0   1   ....

Followup Test 1:    x   x   0   x   x   0   x   1   ....

Followup Test 2:    x   x   0   x   x   1   x   x   ....

MMEL Exceeded?:     N   N   N   N   N   Y   N   Y   ....
```

For the purpose of counting MMEL exceedances, the schema above is equivalent to the set of follow-up sequences shown below, where in the last column in the third row '0/1' indicates that either followup test outcome, 0 or 1, may occur without affecting the MMEL outcome. As a practical matter, the 2nd followup test would not be done, but for the purpose of counting, it is convenient to pretend that two parallel sequences of followup tests did occur. Therefore, to count MMEL exceedances, one can randomly generate three parallel series of TST outcomes as shown, and classify MMEL outcomes as follows: MMEL is exceeded only if (a) the first row contains 1 and (b) either the first or second row contains 1. This makes it very easy to classify MMEL outcomes for a sequence of scheduled routine monitoring tests, generating the fourth row. (This describes one way to count MMEL exceedances using simulated TST outcomes. It is also possible to do so using direct probability calculations - see Appendix 2).

```
Scheduled Tests:    0   0   1   0   0   1   0   1   ....

Followup Test 1:    x   x   0   x   x   0   x   1   ...

Followup Test 2:    x   x   0   x   x   1   x   0/1 ....

MMEL Exceeded?:     N   N   N   N   N   Y   N   Y   ....
```

Each sequence of 0s and 1s can be simulated by using a function that generates random samples from a binomial distribution with N=1 (i.e., the Bernoulli distribution) using the probability of a TST 'fail'. That can be done in statistical software and in spreadsheets. The last row is produced by a logical ('IF') function.

One may also be interested in the expected frequency of one versus two followup tests. These are readily calculated directly from the sequences above.

The next step is to determine how often MMEL exceedances occur back-to-back (that is, in 'runs' of 2 or more successive exceedances). With reference to a simulated sequence of MMEL exceedances and non-exceedances, this can be framed as follows. If we select any schedule monitoring test (i.e, any position in the sequence above), and also examine the scheduled test following it, are they both MMEL exceedances? (Examining both the preceding and next test would amount to 'double counting'). Another way to frame this is to suppose that monitoring has just begun: what is the probability the first two scheduled tests result in MMEL exceedances? Then think of repeating this experiment a million times.

Appendix 2: Concepts for Probability Calculations

All calculations assume outcomes of all WET tests are independent of one another and that the probability of failing TST remains constant.

Probability of Failing TST

The probability of a TST 'fail' (failure to reject the null hypothesis) is made using the non-central t distribution, for specified CV and percent effect. The non-central t calculations can be done in a spreadsheet. The key is to correctly specify the noncentrality parameter value as a function of CVs (or standard deviations) and percent effect. My calculations used R.

Tables of expected probabilities for TST fails are provided separately in spreadsheets named "ptst.n102030.sdconst.csv" and "ptst.n102030.cvconst.csv". The first used the same standard deviation (SD) for controls and water samples (RWC). The second used the same coefficient of variation (CV) for the two samples. These two alternative approaches were discussed by Fox et al. (2019) and the choice between them should have little influence on results when percent effect is 10% or less. Calculations are based on power calculations for the t-test, using the non-central t-distribution. These were generated using the R function "TST.pwr.fn2" (R functions are provided in a text file,"R-functions.txt").

Probability of Exceeding MMEL

The probability that a WET test fails TST will be symbolized *p*. Conditional on such a 'fail', and assuming that this probability does not change, the probability that the

first or second followup test also fails is given by[1] $p + (1-p)*p$. Thus, the probability of an MMEL exceedance is $P_E = p*(p + (1-p)*p)$. (Notice that this does not refer to runs or multiple MMEL exceedances, it refers to the probability that any one scheduled monitoring test leads to an MMEL exceedance). The probability of back-to-back MMEL exceedances is P_E^2 (refer to the conceptual experiment at the end of Appendix 1). The tables below were calculated using simulated sequences and the equations described above.

Table A1-1. Probabilities of MMEL Exceedances.

	Probability MMEL exceeded		Probability of back-to-back exceedances	
Probability of TST-fail	Simulated	$P_E = p*(p+(1-p)*p)$	Simulated	P_E^2
0.02	0.000759	0.000792	0	6.27e-07
0.04	0.00316	0.00314	8e-06	9.83e-06
0.06	0.00702	0.00698	6.3e-05	4.88e-05
0.08	0.0124	0.0123	0.000177	0.000151
0.10	0.0191	0.0190	0.000389	0.000361
0.12	0.0271	0.0271	0.000737	0.000733
0.14	0.0366	0.0365	0.00132	0.00133
0.16	0.0473	0.0471	0.00223	0.00222
0.18	0.0592	0.0590	0.00353	0.00348
0.20	0.0723	0.0720	0.00521	0.00518
0.22	0.0865	0.0862	0.00754	0.00742
0.24	0.102	0.101	0.0104	0.0103
0.26	0.118	0.118	0.0140	0.0138
0.28	0.135	0.135	0.0183	0.0182
0.30	0.153	0.153	0.0236	0.0234
0.32	0.172	0.172	0.0297	0.0296
0.34	0.192	0.192	0.0370	0.0368
0.36	0.213	0.213	0.0455	0.0452
0.38	0.235	0.234	0.0550	0.0547
0.40	0.257	0.256	0.0659	0.0655

[1] This could be modified to $p_1 + (1-p_1)*p_2$, assigning different probabilities to the first and second followup tests.

Appendix 3: Calculating the Probability that an MMEL Exceedance Occurs in a Sequence of WET Tests

The probabilities that at least one of T scheduled tests[2] results in a MMEL exceedance or that no exceedance occurs are based on binomial probability calculations using P_E above. Figure 1 was constructed by calculating P_E as above and then calculating the appropriate binomial probability. Those calculations were made in R, but can also be made in a spreadsheet,.

Using the Excel® function as an example, the binomial probability that at least one exceedance occurs in a sequence of T scheduled monitoring events is

1 - BINOM.DIST(0, T, P_E, FALSE)

This statement evaluates the probability that no exceedance is observed in T events (binomial trials), and subtracts it from 1, giving the binomial probability that 1 or more exceedances are observed.

The statement "= BINOM.DIST(0, T, P_E, FALSE)" will give the probability that no exceedances occur.

[2] To be absolutely clear: "... at least one of T scheduled WET tests 'fails' TST and then is followed by one or two followup tests, resulting in a MMEL exceedance ..."

References

Fox, JF, DL Denton, J Diamond and R Stuber. 2019. Comparison of False-Positive Rates of 2 Hypothesis-Test Approaches in Relation to Laboratory Toxicity Test Performance. Environmental Toxicology and Chemistry 38(3):511-523. DOI:10.1002/etc.4347

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

US Environmental Protection Agency. 2002. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. Third edition. EPA/821/R-02/013. Office of Water, Washington, DC

US Environmental Protection Agency. 2010. National pollutant discharge elimination system test of significant toxicity technical document. EPA/833-R-10-004. Office of Wastewater Management, Washington, DC.

Zheng L, Diamond J, Denton DL. 2013. Evaluation of whole effluent toxicity data characteristics and use of Welch's T-test in the test of significant toxicity analysis. Environ Toxicol Chem 32:468–474

Reported in August 2019 (in black)
New calculations December 2019 (red, bold)

Table 2. Probability of at least one run of 2 or more MMEL failures		
Probability *p* of failing TST during routine monitoring	P_V[1], probability of MMEL violation each month	Probability of one or more runs (2 successive MMEL failures) in 60 months, resulting in TRE (based on P_V)[2]
0.04	**0.003136**	**9.83E-06**
0.05	0.004875	0.0011
0.06	**0.006984**	**4.88E-05**
0.10	0.019000	0.020
0.10	**.019000**	**0.000361**
0.12	0.027072	0.042
0.12	**0.027072**	**0.000733**
0.13	0.031603	0.06
0.14	**0.036456**	**0.001329**
0.14	**0.036456**	**0.001329**
0.15	0.041625	0.0930
0.16	**0.047104**	**0.002219**
0.20	0.072000	0.25
0.20	**0.0720**	**0.0052**
0.24	**0.1014**	**0.0103**
0.25	0.109375	0.48
0.26	**0.1176**	**0.0138**

[1] $P_V = p*p + p*(1-p)*p$
[2] Average of six simulation runs, each consisting of 10,000 simulated sequences of 60 months