CALIFORNIA
Water Boards

Lahontan Regional Water Quality Control Board

April 12, 2012

Via email

Harold J. Singer
Executive Officer
Lahontan Regional Water Quality Control Board

COMMENTS ON CONSIDERATION OF AMENDMENT OF ORDER NO. R6V-2011-0005A1 (ORDER) ISSUED TO PACIFIC GAS AND ELECTRIC COMPANY (PG&E)

This letter provides comments, solicited by your March 22, 2012 letter on the subject line Order. The Order at issue requires, in section 3.a., that PG&E propose a method or methods to perform an initial and quarterly evaluation of every domestic or community well in the affected area to determine if detectable levels of hexavalent chromium between the maximum background and the Public Health Goal concentrations represent background conditions, or are more likely than not, partially or completely, caused by the discharge of waste by PG&E. The maximum background level for hexavalent chromium in groundwater in the Hinkley Valley is 3.1 parts per billion (ppb); the Public Health Goal for hexavalent chromium in drinking water is 0.02 ppb.

PG&E has responded by concluding that developing a methodology as required in section 3.a. of the Order is not feasible. In its December 22, 2011 letter, PG&E outlined potentially feasible statistical methods to comply with the Order, and concluded they were invalid for several reasons. Water Board staff disagree with PG&E's reasons for its conclusion, and believe that at least two of the statistical methods discussed by PG&E would meet the requirements of section 3.a. of the Order.

On April 10, 2012, PG&E submitted the *Replacement Water Feasibility Study*, in response to section 2.c. of the Order. The Prosecution Team is currently reviewing that submittal for compliance with the Order, and will provide its comments under separate cover. This letter provides comments only on the subject of feasible statistical methods, as requested in your March 22, 2012 letter. Below is a summary of PG&E's reasons for dismissing available statistical methods in **bold font**, followed by Water Board staff's rebuttal.

1. **Using a different statistical test (in addition to the upper tolerance limit (UTL) method that was used to determine the maximum background value for hexavalent chromium) would lead to excessive inflation of the overall false positive rate.**

The Cleanup and Abatement Order requires that the Discharger propose a method or methods to perform an initial and quarterly evaluation of every domestic or community well in the affected area. When considering trends in chromium rates at individual wells, the Discharger must take into consideration both the existing UTL and additional statistical analysis described in Finding 26 of the Order. Thus, the CAO requires PG&E to submit a method, in the form of a different statistical test in addition to the UTL, to determine if an increasing trend is present in a single given well, indicating, more likely than not, the well is impacted by PG&E's waste chromium. A requirement to supply replacement water may be triggered by either an exceedance of the UTL, or use of a different statistical test.

The concern with combining two or more statistical tests to determine a trend in chromium levels is that the false positive (or Type I) error will result in *wrongly* concluding that an increasing trend of hexavalent chromium in a well indicates it is impacted by PG&E's discharge of waste chromium. Water Board staff acknowledge that using a different statistical test (in addition to the UTL) will increase the false positive rate. However, we disagree that the increase is excessive. Since either test can be used as the basis for ordering replacement water for an individual well, the overall false positive rate is, at most, the sum of both false positive rates for each individual statistical test.

An acceptable false positive rate should be chosen considering the hypothesis being tested, the implications of the test outcome, and the amount of risk one is willing to accept in drawing an incorrect conclusion. In statistical hypothesis testing, false positive error rates commonly range from one to ten percent, with five percent (one in twenty) being a typical value. The false positive rate for the UTL used to develop the maximum background level is five percent, and the false positive rate for a one-tailed Spearman's test (for example) is also five percent. Applying both tests would result in a false positive rate of at most ten percent (personal communication, Dr. Neil Willits, University of California statistician, March 29, 2012). This is within the range of standard error rates for groundwater compliance and assessment monitoring (US EPA, 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities Unified Guidance).

Here, the statistical hypothesis being tested is whether a given Hinkley drinking water well has been impacted by PG&E's waste chromium. The CAO requires using two statistical methods to test this hypothesis, which may result in an increase in the false positive error rates. The implications of the test outcome are very significant to the well user, with possible real-world impacts for human health, quality of life, and free use of

property. The Prosecution Team argues that a ten percent false positive rate (one in ten) should be acceptable for the statistical hypothesis being tested in Hinkley.

A ten percent false positive error rate means that one well in ten tested could be deemed "impacted" by waste when it is not. A one in ten chance of wrongly concluding a well is impacted is a reasonable amount of risk when human health is at issue and is consistent with the principles of environmental justice. A conservative approach erring on the side of health protection is justified and appropriate, especially when considering the long-term stress and worry borne by those residents of Hinkley whose wells show increasing levels of a harmful chemical. The burden of the regulatory uncertainty should fall on PG&E, and not the residents of Hinkley.

The Prosecution Team also notes that Title 27 of the California Code of Regulations (CCR), which is cited in Finding 26 of the CAO, provides guidance for acceptable facility-wide false positive rates where retest methods differs from the initial test method: ". . . the discharger shall demonstrate that the initial and retest method, in combination, provide a facility-wide false positive rate of greater or equal to than five percent." *See* 27 CCR § 20415(e)(8)(E)(5)(b). Thus, the allowable facility-wide false positive rate is <u>no less than</u> 5 percent but is not limited in the upper bound of false positive rates. In other words, a five percent risk of a false positive conclusion is the *minimum* amount of risk allowed a discharger by law, not the *maximum* amount.

Lastly, the false positive rate is balanced by Finding 26 of the Order. An increasing trend of concentrations provides one piece of evidence to make a determination of an affected well, but is not considered in isolation of hydrogeologic evidence. Finding 26 of the Order requires PG&E to consider a number of factors in its analysis, including but not limited to, changes in hexavalent chromium over time (i.e., trend analysis), location of the well in relationship to the plume and groundwater flow direction, and isotopic analysis of hexavalent chromium. These additional considerations allow the statistical analysis to be used as a primary screening tool, against which additional evidence is examined.

For the above reasons, the Prosecution Team **disagrees** with PG&E's conclusion that the application of a statistical test to determine trends in a given well would *excessively* inflate the false positive rate.

2. **Feasible statistical methods require sample sizes of at least 8 or greater. Less than 10 percent of relevant wells have this number of samples, so none of the feasible methods are applicable to determine both statistically and environmentally significant trends.**

The numbers of samples needed depends on the strength of the trend. Increasing trends can be detected at a significant level with four samples (personal communication, Dr. Neil Willits). Therefore, feasible statistical methods to evaluate trends, such as Spearman's or Man-Kendall correlation test, can be conducted with a minimum of four samples.

The Prosecution Team does not agree with PG&E's conclusion that because the current dataset for some wells may not be large enough to evaluate, that no statistical test is feasible for all drinking water wells. While a well may not have the minimum number of samples needed *today*, such a dataset could be generated within a reasonable amount of time through PG&E's well sampling program, residents' sampling results from an independent third party, or Water Board staff sampling. The absence of a valid current dataset does not preclude the generation of a valid future dataset from which to apply appropriate trend analysis methods.

3. **Feasible tests (Mann-Kendall, Spearman's) do not distinguish environmental from statistical significance. The significance of any statistical trend needs to be evaluated in the context of hydrogeology, adjacent well data, remediation activities, and seasonality.**

As discussed in number 1 above, Finding 26 of Order No. R6V-2011-0005A1 allows consideration of a number of factors to judge hydrogeological and environmental factors along with a statistically significant test result. The statistical method is a screening tool, against which environmental evidence and common sense are applied. Therefore, this argument does not support PG&E's contention that no feasible method exists.

Conclusion

The Prosecution Team disagrees with PG&E's evaluation of feasibility of available statistical methods. As described above, we disagree:

1) That the false positive rate associated with using two statistical tests is excessive;
2) That the number of samples needed presents an insurmountable obstacle to conducting the required analysis; and
3) That conducting the required analysis restricts or precludes consideration of the environmental significance of the test results.

The Prosecution Team believes that at least two statistical tests mentioned in PG&E's response are feasible to fulfill the requirements of section 3.a. of the Order: Mann-Kendall or Spearman's test for intra-well trend analysis, according to the following specifications:

> Mann-Kendall: Data from each individual well will be tested for increasing trends between the sampling date and the measured chromium concentration using the Mann-Kendall test. Tests will be run as 1-tailed tests. The conclusion that an increasing trend is present will be accepted whenever the p-value is less than a 2-tailed alpha value of 0.10 and tau or S is greater than 0. The focus is only on positive values of tau or S, since those are the only ones that are indicative of an increasing trend. For this reason, a 2-tailed alpha value of .10 will correspond to a standard one-tailed alpha value of 0.05. For the purposes of this analysis,

values below the detection limit will be treated as if they are equal to that limit. In the event the detection limit changes, all non-detect values will be considered numerically equal.

Spearman's rho: Data from each individual well will be tested for increasing trends between the sampling date and the measured chromium concentration using the Spearman's Rank Correlation Coefficient. Tests will be run as 1-tailed tests. The conclusion that an increasing trend is present will be accepted whenever the 2-tailed p-value is less than an alpha value of 0.10 and rho is greater than 0. The focus is only on positive values of rho, since those are the only ones that are indicative of an increasing trend. For this reason, a 2-tailed alpha value of .10 will correspond to a standard one-tailed alpha value of 0.05. For the purposes of this analysis, values below the detection limit will be treated as if they are equal to that limit. In the event the detection limit changes, all non-detect values will be considered numerically equal.

The minimum dataset required per well should be four samples. If third party sampling (i.e., sampling by Water Board staff or by an independent consultant/laboratory hired by a well owner) is conducted in order to generate the minimum dataset for any well in the affected area, staff proposes the following guidelines.

Frequency of Sampling: Groundwater sampling of domestic or community supply wells should be conducted at least 2 weeks apart, to ensure that sample results represent a hydrogeologically-independent sample of groundwater (US EPA, 2009).

Sample Collection and Analysis: Sample collection should be conducted by a qualified professional under proper chain-of-custody procedures. Sample analysis should be conducted at a California Environmental Laboratory Accreditation Program (ELAP) certified lab.

Sample Reporting: Chain of custody sheets, laboratory data sheets, quality assurance and quality control data from the lab, and sampler qualifications must be included with sampling results.

Once the minimum dataset for any well in the affected area is generated, PG&E should conduct the statistical evaluation on a quarterly basis using one of the two tests noted above, to assess if an increasing trend is present. If such a trend is present, the well in question should be considered an impacted well, unless PG&E provides evidence as described in Finding 26 of Order No. R6V-2011-0005A1. The Executive Officer must concur in writing with the evidence presented by PG&E that the well is not impacted.

Thank you for the opportunity to provide input on this issue. Please contact me at 530-542-5436 with any questions.

Lauri Kemper, PE
Assistant Executive Officer

cc: Dr. Neil Willits, UC Davis Statistical Laboratory

ALH/adw/T: WHW CAO amendment comments.docx
File Under: 6B369107001