

INTEROFFICE CORRESPONDENCE

DATE: November 3, 1993

TO: G. A. Anderson, Environmental Engineering and Technology, Bldg. 080, X8505

FROM: M. A. Siders, Geosciences, Bldg. 080, X6933 *MAS*

SUBJECT: STATISTICAL METHODOLOGY FOR SELECTION OF POTENTIAL CONTAMINANTS OF CONCERN (PCOCS) - MAS-017-93

PURPOSE

This letter is to call attention to an issue that will have substantial impacts on all future work regarding the selection of PCOCs in all Operable Units (OUs). This issue is the statistical methodology proposed by EG&G for statistical comparisons of OU data versus background data.

BACKGROUND

For the past nine months, I have been involved with the issue of OU versus background comparisons. I have attended numerous meetings with the Colorado Department of Health (CDH) and the Environmental Protection Agency (EPA) which failed to produce a consensus on the statistical methodology that should be used for the comparison. Due to this stalemate between the Department of Energy (DOE)/EG&G and the regulatory agencies, Dr. Dick Gilbert (Pacific Northwest Labs) was brought in as a statistical expert and "arbitrator" for the dispute. All parties involved agreed to follow the methodology that Dr. Gilbert agreed to propose in a report to EPA.

Gilbert's report was released on July 30, 1993; his recommendations were lengthy and covered a battery of statistical tests that could be applied for data comparisons. Dr. Gilbert was careful to discuss why some tests were "better" (i.e., more powerful) than others for a given data distribution or for a given distributional shift from background. Because of the level of detail involved in Gilbert's report, the regulatory agencies requested that DOE/EG&G provide them with a specific proposal as to how Dr. Gilbert's recommendations would be implemented.

DISCUSSION

Over the past several months, I worked with Dennis Smith and Rick Roberts to put together a proposed methodology for PCOC selection. On September 29, 1993, this methodology was presented to DOE, EPA, and CDH. The flow chart distributed at this meeting clearly indicated that either the Gehan test or nonparametric ANOVA tests would be used for the comparisons (see Attachment A). However, the EPA's official comments on EG&G's proposed methodology (received October 25, 1993) contained a different flow chart, with *only* the Gehan test listed (see Attachment B).

DOCUMENT CLASSIFICATION
REVIEW WAIVER PER
CLASSIFICATION OFFICE

G. A. Anderson
November 3, 1993
MAS-017-93
Page 2

The present concern – which is documented in two memos sent to Dennis Smith (dated September 2 and September 14) – is that EG&G is about to lock into a statistical methodology (the Gehan test) that is wrought with potential problems. These problems with the Gehan test are of both a practical and a technical nature.

The Gehan test was proposed by the Statistical Applications Group (SA) on plantsite who were acting as the EG&G/DOE statistical experts because of the present lack of statistical expertise within Environmental Restoration Management (ERM). Unfortunately, these statisticians have little understanding of the geological, geochemical, and hydrological aspects that strongly influence our work. Neither do they seem to have much feel for the more practical aspects of managing subcontractors (who do much of the OU work), or for the costs and delays that may arise by implementing an obscure statistical test for OU versus background comparisons.

Our first memo to Dennis Smith (MAS-006-93, Attachment C) reviewed the context of the Gilbert report, including the practicality of the statistical tests that Gilbert discussed in his report. At this time, we cautioned against use of the Gehan test, of which Gilbert himself said "As the performance of the Gehan test has not, in my opinion, been adequately determined, I recommend that statistical evaluations and comparisons of its performance with competing tests should be conducted by EG&G at the earliest time." Competing tests include the Wilcoxon Rank Sum and Kruskal-Wallis tests, which, according to Gilbert "...are very well known by statisticians and practitioners, and are widely used in many fields of application" (Attachment D). As far as I know, this comparative testing of the Gehan test has not been done.

Our second memo to Dennis Smith (MAS-007-93, Attachment E) states more firmly our reticence to lock ourselves into an untested methodology (i.e., the Gehan test), without the option of using other nonparametric tests. EPA and CDH would be perfectly happy to let us use the standard nonparametric ANOVA tests (i.e., Wilcoxon and Kruskal-Wallis), but the Gehan ball is rolling along unimpeded, and threatening to flatten all other viable options.

RECOMMENDATIONS

EG&G's recommended policy should incorporate practicality as well as technical correctness; the Gehan test is not practical, and its technical correctness has been questioned in the literature (Helsel, 1990, *ES&T*; Attachment F). Schedule is also of concern; SA recently did some work for OU 11 in which they took 200 hours to do the OU versus background comparison using the Gehan test. In contrast, statisticians using standard nonparametric ANOVA tests should have completed the work in less than 40 hours. The Gehan method was developed by SA. Cost and schedule impacts from implementation by subcontract staff would be equally or more significant, because the Gehan test – and instructions on its use – are not available in any statistical software package.

G. A. Anderson
November 3, 1993
MAS-017-93
Page 3

By implementing untested statistical methods, EG&G is setting itself up for failure. Flexibility in this area is crucial, and a lack of recognition of this by EG&G statistical "experts" is demonstration of a general lack of understanding of the ERM program and goals. The support offered by SA is valuable from a technical basis but should be strictly limited to their area of expertise. ERM cannot afford to recommend policy to DOE without carefully integrating the technical recommendations from every contributor who has supported this effort with ERM program goals. A lack of foresight at this critical stage could have serious impacts on future work.

It is strongly recommended that the statistical methodology proposed by EG&G to DOE should include the option of using the widely accepted, standard nonparametric ANOVA tests (i.e, the Wilcoxon Rank Sum and Kruskal-Wallis).

jlm

Attachments:
As Stated (5)

cc:
C. A. Bicher
W. S. Busby
M. E. Levin
T. P. O'Rourke
R. S. Roberts
S. H. Singer

BACKGROUND COMPARISON METHODOLOGY

