From: POULSEN Mike
To: Burt Shephard/R10/USEPA/US@EPA
Cc: Eric Blischke/R10/USEPA/US@EPA
Subject: RE: Risk Parameter Table and Upstream Chemical Concentrations - 3rd time is the charm
Date: 02/13/2007 09:43 AM

Burt -

You are certainly correct that the 95th percentile of a normal distribution is mean + 1.645*SD, or about 1.7*SD if you correct for number of samples. Eric approximated this by 2*SD. A little rough, but an easier calculation to do in your head. I see now that in Chris' refinement, he looked at two-tails instead of one. We'll have to fix that when we get around to doing a rigorous evaluation of the data. But by then I'll advocate using the 90th percentile, or approximately mean + 1.28*SD (if the data are normally distributed). That's what Washington does, and that's what I've done in the past in Oregon. I don't think I'll advocate tolerance limits. However, we haven't yet agreed that these are the data we want to evaluate. I just wanted to let people know that we need to look beyond the mean in evaluating upstream or background data. I'm looking forward to digging into upstream/background data in the future. And I'll need to look more closely at EPA guidance when we get to that point.

- Mike

-----Original Message-----
From: Shephard.Burt@epamail.epa.gov
[mailto:Shephard.Burt@epamail.epa.gov]
Sent: Monday, February 12, 2007 4:51 PM
To: POULSEN Mike
Cc: Blischke.Eric@epamail.epa.gov
Subject: RE: Risk Parameter Table and Upstream Chemical Concentrations - 3rd time is the charm

Mike,

It may have been Eric's intent to calculate the upper 90th or upper 95th percentile of the upstream data, but that's not what his spreadsheet does, nor is it what his original e-mail indicated he was trying to do. Its true that the formula for estimating the xth percentile (xp) of a normally distributed data set is:

xp = mean + (Zp x std. deviation)

which is a fancy way of saying Zp is the pth percentile of the standard normal distribution. If Zp = 2, the mean + (2 x s) yields the 97.72th percentile of the data set, based on Zp values for a cumulative normal distribution in Table A1 of Gilbert (1987). This was Eric's original spreadsheet. Chris Thompson and I both interpreted Eric's original e-mail and spreadsheet as an attempt to calculate the 95% upper confidence limit of the mean, not the 95th percentile of a data set, which are two very different summary statistics. Assuming a normally distributed data set, the 90th and 95th percentiles of any given normally distributed data set are calculated as:

x0.95 = mean + (1.645 x std. deviation) x0.90 = mean + (1.28 x std. deviation)

Lesson to be learned, we need to be more precise and specific as to our objective during any efforts we make of statistically evaluating site data from Portland Harbor.

Best regards,

Burt Shephard
Risk Evaluation Unit
Office of Environmental Assessment (OEA-095)
U.S. Environmental Protection Agency, Region 10
1200 6th Avenue
Seattle, WA 98101

Telephone: (206) 553-6359
Fax: (206) 553-0119

e-mail: Shephard.Burt@epa.gov

POULSEN Mike
<POULSEN.Mike@deq.state.or.us>

02/12/2007 03:01 PM

To
Burt Shephard/R10/USEPA/US@EPA
cc
Eric Blischke/R10/USEPA/US@EPA
Subject
RE: Risk Parameter Table and Upstream Chemical Concentrations - 3rd time is the charm

Burt -

At the data retreat, I suggested that we look at the upper percentile of the upstream data, not an upper confidence limit on the arithmetic mean of the upstream data. No need to bring in the standard error. I took this suggestion from the approach used by Washington DOE. My understanding is that DOE looks at the upper 90th percentile of background data. A UCL on the arithmetic mean of site data can then be compared with the upper percentile of background data to determine if the site is above background. DOE has other tests, including tolerance limits, but they are for complying with other DOE requirements (e.g., a 95% confidence interval around the 90th percentile), and I don't think we need to go there right now.

Eric took a quick look at the PH upstream data by calculating the 95th percentile assuming a normal distribution. This is roughly the mean plus 2 SDs. I agree with Eric that this is an OK first cut. Chris refined the calculation a bit using the actual number of data points, and got essentially the same result. I haven't even looked to see if the data are normally distributed. We can use a non-parametric estimation of the upper percentile if we need to. At some point, I'm sure we'll want to dig deeper into the evaluation of upstream data and refine our calculations. Perhaps this can be a future TCT topic. I'm not sure how much additional effort we need right now.

- Mike

-----Original Message-----
From: Shephard.Burt@epamail.epa.gov
[mailto:Shephard.Burt@epamail.epa.gov]
Sent: Monday, February 12, 2007 12:42 PM
To: Chris Thompson
Cc: ANDERSON Jim M; Aron Borok; audiehuber@ctuir.com; BBarquin@hk-law.com; Humphrey.Chip@epa.gov; Cinde Donoghue; csmith@parametrix.com; Black.Curt@epa.gov; Davoli.Dana@epa.gov; Blischke.Eric@epa.gov; eri GAINER Tom; Grepo-Grove.Gina@epa.gov; (b) (6) howp@critfc.org; jay.field@noaa.gov; jeff.b ; Jennifer Arthur; jeremy_buck@fws.gov; Goulet.Joe@epa.gov; Smith.Judy@epa.gov; TOEPEL Kathryn; Koch.Kristine@epa.gov; Lisa.Bluelake@grandronde.org; Cora.Lori@epa.gov; Ader.Mark@epa.gov; MCCLINCY Matt; Parker Wittman; PETERSON Jenn L; Pam Bridgen; POULSEN Mike; Fuentes.Rene@epa.gov; rgensemer@parametrix.com; Robert.Neely@noaa.gov; Ron.Gouguet@noaa.gov; rose@yakama.com; Sheldrake.Sean@epa.gov; sheila@ridolfi.com; tomd@ctsi.nsn.us; Valerie Lee
Subject: RE: Risk Parameter Table and Upstream Chemical Concentrations - 3rd time is the charm

Chris et al.

Attached is yet another revision of the upstream chemical concentration spreadsheet. I've provided a workaround for a "feature" of Excel that for some reason won't permit accurate estimation of confidence intervals without a little bit of work on formulas in the cells where confidence limits are estimated. I've only revised the data page of the spreadsheet (the upstreamdatasummary tab of the spreadsheet), I haven't done anything to the various graphs

The problem is that Excel does not contain a built in function to calculate standard error (SE), but only has a function to calculate a sample standard deviation. The formula to calculate a 95% upper confidence limit (assuming the data are normally distributed, an assumption that hasn't been tested with the attached data and which adds complexity to the calculation if its needed) is: mean + (t0.05, d.f. x SE). The standard error in Excel has to be calculated as the sample standard deviation divided by the square root of the number of samples. In Excelspeak, that becomes STDEV(datarange) / SQRT(COUNT(datarange)).

This is actually a well known problem with Excel among statisticians, which you can convince yourselves of by entering Microsoft Excel statistical errors into Google sometime. And if any of you were wondering, Excel's CONFIDENCE function doesn't correctly calculate confidence intervals for samples either. CONFIDENCE can work, but only in the rare instances where the true population standard deviation is known.

(See attached file: upstreamdatasummary_revised Take 3.xls)

Best regards,

Burt Shephard
Risk Evaluation Unit
Office of Environmental Assessment (OEA-095)
U.S. Environmental Protection Agency, Region 10
1200 6th Avenue
Seattle, WA 98101

Telephone: (206) 553-6359
Fax: (206) 553-0119

e-mail: Shephard.Burt@epa.gov

Chris Thompson
<chris.thompson@
EILTD.net>

02/09/2007 02:14
PM

To
Eric Blischke/R10/USEPA/US@EPA,
Burt Shephard/R10/USEPA/US@EPA,
Chip Humphrey/R10/USEPA/US@EPA,
Curt Black/R10/USEPA/US@EPA, Dana
Davoli/R10/USEPA/US@EPA, GAINER
Tom <GAINER.Tom@deq.state.or.us>,
Gina
Grepo-Grove/R10/USEPA/US@EPA,
jeff.baker@grandronde.org,
Jennifer L Peterson
<PETERSON.Jenn@deq.state.or.us>,
jeremy_buck@fws.gov,
anderson.jim@deq.state.or.us, Joe
Goulet/R10/USEPA/US@EPA, Judy
Smith/R10/USEPA/US@EPA,
Kathryn.Toepel@state.or.us,
Kristine Koch/R10/USEPA/US@EPA,
MCCLINCY Matt
<MCCLINCY.Matt@deq.state.or.us>,
howp@critfc.org, POULSEN Mike
<POULSEN.Mike@deq.state.or.us>,
Rene Fuentes/R10/USEPA/US@EPA,
Robert.Neely@noaa.gov, Sean
Sheldrake/R10/USEPA/US@EPA,
tomd@ctsi.nsn.us, Parker Wittman
<parker.wittman@EILTD.net>,
csmith@parametrix.com,
rgensemer@parametrix.com,
rose@yakama.com,
erin.madden@gmail.com,
Ron.Gouguet@noaa.gov, Cinde
Donoghue
<cinde.donoghue@EILTD.net>,
jay.field@noaa.gov, Jennifer
Arthur
<jennifer.arthur@EILTD.net>, Aron
Borok <aron.borok@EILTD.net>,
Lori Cora/R10/USEPA/US@EPA, Mark
Ader/R10/USEPA/US@EPA,
BBarquin@hk-law.com,
audiehuber@ctuir.com,
Lisa.Bluelake@grandronde.org, Pam
Bridgen <pj.bridgen@EILTD.net>,
sheila@ridolfi.com

cc
Valerie Lee
net>,
(b) (6) Pam Bridgen
et>, Pam
Bridgen <pj.bridgen@EILTD.net>

Subject
RE: Risk Parameter Table and
Upstream Chemical Concentrations

Hi Eric et al.,

I attached a slightly revised version of the upstream data summary that Eric sent out earlier this afternoon; as Eric noted, the estimated upper 95% confidence limit in the spreadsheet is an estimation, i.e. not completely accurate. I revised the spreadsheet to incorporate the correct t-values for the corresponding sample size of each analyte, and recalculated the upper 95% confidence limit of the individual data values presented in the spreadsheet. Except for chlordane (sample size = 3), the correction caused only minor increases in the 95% estimate because most sample sizes were about 30 (20-39).

Cheers,

Chris

Chris Thompson, Ph.D.
Environment International, Ltd.
5505 34th Ave NE
Seattle, WA 98105
(206) 525-3362
(206) 525-0869 (fax)

THE INFORMATION CONTAINED IN THIS E-MAIL MESSAGE IS LEGALLY PRIVILEGED AND CONFIDENTIAL INFORMATION INTENDED SOLELY FOR THE USE OF THE PERSONS OR ENTITIES NAMED ABOVE. IF YOU ARE NOT SUCH PERSONS OR ENTITIES, YOU ARE HEREBY NOTIFIED THAT ANY DISTRIBUTION, DISSEMINATION OR REPRODUCTION OF THIS E-MAIL MESSAGE IS STRICTLY PROHIBITED. IF YOU HAVE RECEIVED

THIS MESSAGE IN ERROR, PLEASE IMMEDIATELY CALL US AT 206-525-3362.

THE INFORMATION CONTAINED IN THIS E-MAIL MESSAGE IS CONFIDENTIAL INFORMATION INTENDED SOLELY FOR THE USE OF PERSONS OR ENTITIES NAMED ABOVE. IF YOU ARE NOT SUCH PERSONS OR ENTITIES, YOU ARE HEREBY NOTIFIED THAT ANY DISTRIBUTION, DISSEMINATION OR REPRODUCTION OF THIS E-MAIL MESSAGE IS STRICTLY PROHIBITED. IF YOU HAVE RECEIVED THIS MESSAGE IN ERROR, PLEASE IMMEDIATELY CALL US AT 206-525-3362.

-----Original Message-----
From: Blischke.Eric@epamail.epa.gov
[mailto:Blischke.Eric@epamail.epa.gov]
Sent: Friday, February 09, 2007 1:01 PM
To: Shephard.Burt@epamail.epa.gov; Humphrey.Chip@epamail.epa.gov; Black.Curt@epamail.epa.gov; Davoli.Dana@epamail.epa.gov; GAINER Tom; Grepo-Grove.Gina@epamail.epa.gov; jeff.baker@grandronde.org; Jennifer L Peterson; jeremy_buck@fws.gov; anderson.jim@deq.state.or.us; Goulet.Joe@epamail.epa.gov; Smith.Judy@epamail.epa.gov; Kathryn.Toepel@state.or.us; Koch.Kristine@epamail.epa.gov; MCCLINCY Matt; howp@critfc.org; POULSEN Mike; Fuentes.Rene@epamail.epa.gov; Robert.Neely@noaa.gov; Sheldrake.Sean@epamail.epa.gov; tomd@ctsi.nsn.us; Parker Wittman; csmith@parametrix.com; rgensemer@parametrix.com; rose@yakama.com; erin.madden@gmail.com; Ron.Gouguet@noaa.gov; Cinde Donoghue; jay.field@noaa.gov; Jennifer Arthur; Chris Thompson; Aron Borok; Cora.Lori@epamail.epa.gov; Ader.Mark@epamail.epa.gov; BBarquin@hk-law.com; audiehuber@ctuir.com; Lisa.Bluelake@grandronde.org; Pam Bridgen; sheila@ridolfi.com
Subject: Risk Parameter Table and Upstream Chemical Concentrations

Attached are a couple of spreadsheets that were discussed at the data retreat.

The first is the risk parameter table. This table includes all the screening values used in the presentations. Sources of information are provided in the notes sections. The correct TEC value for chlordane has been inserted (the previous number was the NOAA PEL).

The second is a summary of upstream data for selected chemicals. This data was extracted from QM. Only data from between Ross Island and Willamette Falls are included. Some charts are also included which present the minimum, mean, maximum and an estimate of the 95% of the distribution. The 95% distribution was estimated through the following calculation: mean + 2(standard deviation). While this estimate is not statistically rigorous, it does provide some information.

If you have any questions, please let me know and I will try to answer them.

Thanks, Eric

(See attached file: RiskParameters020907.xls)(See attached file: upstreamdatasummary.xls)
[attachment "upstreamdatasummary_revised.xls" deleted by Burt Shephard/R10/USEPA/US]