I. Sampling Data Quality Review

All scripts can be found on the NE Port Biological Sampling Program wiki page (https://www.st.nmfs.noaa.gov/confluence/display/PBSP/NE+Port+Biological+Sampling+Program) under Program Management/Quality management/Scripts.

A. Use the BSMDBS Audit SQL script to compile sampling data from the Biological Sample Monitoring Database System (BSMDBS) data tables and copy and paste into an excel spreadsheet.

B. Add a column for staff type and fill in as appropriate (FTE, Contract).

C. Use the various excel features to check for potential data issues. Indicate what the potential issue is in the ISSUE 1 or ISSUE 2 column. If more than two issues are found for a single row of data, add additional issue columns. Any observations that may help resolve the issue should be added to the ADDITIONAL COMMENTS column.

1. Potential issues include:

a) Duplicate DOCNs: Repeated DOCNs within the same month

b) Duplicate samples: Samples with the same PERMIT, NESPP4, and DATELAND

c) Incorrect species: Species that are not typically sampled or invalid NESPP4 codes

d) Incorrect market: NESPP4 does not match to a valid market

e) No sampler information

f) Samples with the same VTRSERNO, but different vessels

g) Mismatched permit: The PERMIT entered does not match with VESSELNAME entered

h) Missing area: No AREA information entered

i) Missing VTRSERNO: No VTRSERNO information entered (exception hagfish from state vessels)

j) Missing AREA and VTRSERNO: no area or VTRSERNO information entered

k) Incorrect VTRSERNO: VTRSERNO does not fit typical VTRSERNO format (e.g. number of digits)

l) Missing REGION: no region information entered

m) Incorrect GEAR: invalid gear code entered

n) Missing WEIGHT: no weight information entered (where applicable)

2. Use the conditional formatting function to highlight duplicates (one column at a time) in the YEARMONTHDOCN, NESPP4VTRSENO, NESPP4PERMIT DATELAND, and VTRSERNO columns to check for duplicate DOCNs, duplicate samples, samples with the same VTRSERNO, but different vessels, and mismatched permits.

3. Use the filter feature (one column at a time) on the SPPNM, MKTNM, AREA, VTRSERNO, GEARNM, WT, REGION, and SAMP1 columns to look for missing or potentially incorrect information (where applicable) in these columns.

D. After all issues have been identified, create pivot tables to summarize the number of issues by issue type and staff type.

1. Create a table for each ISSUE column.

a) Row labels: ISSUE, STAFF

b) Values: Count DOCN

2. Combine all of the tables to get the total number of each issue.

E. Create a spreadsheet of the identified issues for each staff type.

1. Each issue type should be in a separate tab within each spreadsheet.

2. Include a summary tab with the issue totals (from D above) for the indicated staff type.

3. Verify that the number of rows in each tab matches the summary totals.

F. Post the finished spreadsheets to the Data Quality Review Results page of the NE Port Biological Sampling Program Wiki page (*https://www.st.nmfs.noaa.gov/confluence/pages/viewpageattachments.action?pageId=43287659*).

G. Issues will be forwarded to appropriate staff for correction or comment as follows:

1. FTE: Create a JIRA in the NEPBSP project with FTE spreadsheet created in E above attached for each staff member that has more than one issue identified. Add the Section Manager and COR as watchers.

2. Contractor: COR will create a JIRA in the NEPBSP project to forward issues to the contractor.

II. Samples Requested, Collected and Landings Data Comparison

All scripts can be found on the NE Port Biological Sampling Program wiki page (https://www.st.nmfs.noaa.gov/confluence/display/PBSP/NE+Port+Biological+Sampling+Program) under Program Management/Quality management/Scripts.

A. Use the Requests, Samples, Landings merge spreadsheet to compile samples requested, samples collected, and landings data for a specified sampling quarter.

1. Use the current sampling requests to determine the number of samples, lengths, and ages requested.

   a) Copy and paste above information into the Data tab of the Requests, Samples, Landings merge spreadsheet.

   b) Delete rows where # SAMPLES Requested is zero.

2. Compile sampling data from using the Sample Data SQL script.

   a) Specify the year and quarter in the upper portion of the script, create the table, and run the section of the script that assigns the appropriate REGIONs, AREA_REQs, and GEAR_REQs to the data.

      Before running this section, verify that the correct GEAR_REQ update for mackerel is used (it varies by quarter).

   b) Check the updated table for null values in the NESPP4_GRP and GEAR_REQ fields and AREA_REQ equal to AREA_REQ. Manually add the necessary information to the table and save the finalized data to a spreadsheet for future reference.

   c) Run the last section of the script to total the samples by all requested categories and copy and paste the results into the Data tab of the Requests, Samples, Landings merge spreadsheet.

d) Download and save the current Anticipated Area spreadsheet from the Google drive. Use it to add skate sample information to the data tab of the Requests, Samples, Landings merge spreadsheet.

3. Compile landings data.

a) Use the Clams SQL script to gather surf clam and ocean quahog landings data.

(1) Specify the time period in the upper section of the script where it says A.Cd between…, create the table, and run the section of the script that assigns the appropriate NESPP3, NESPP4, NESPP4_GRP, SPECIES, REGION, AREA_REQ, and GEAR_REQ to the data.

(2) Check the updated table for AREA_REQ equal to AREA_REQ. Manually add the necessary information to the table and save the data in a spreadsheet for future calculations.

(3) Run the second to last portion of the script to total the landings by all requested categories and copy and paste the information into the Data tab of the Requests, Samples, Landings merge spreadsheet.

(4) Run the last section of the script to total the landings by species and market category and copy and paste the results into the Data tab of the Requests, Samples, Landings merge spreadsheet.

b) Use the Landings Data SQL script to gather the other requested species' landings data from the CFDERS tables.

(1) Specify the year, months, and NESPP4s of the requested species and markets in the upper section of the script, create the table, and run the section of the script that assigns the appropriate REGION, AREA_REQ, and GEAR_REQ to the data.

Before running this section, verify that the correct GEAR_REQ update for mackerel is used (it varies by quarter).

(2) Check the updated table for AREA_REQ that equal AREA_REQ. Manually add the necessary information to the table and save the data in to same spreadsheet as the clam landing data (3 a 2 above).

(3) Run the second to last portion of the script to total the landings by all requested categories and copy and paste the information into the Data tab of the Requests, Samples, Landings merge spreadsheet.

(4) Run the last section of the script to total the landings by species and market category and copy and paste the results into the data tab of the Requests, Samples, Landings merge spreadsheet.

4. Use the compiled data to create the following pivot tables in the appropriate tabs of the Requests, Samples, Landings merge spreadsheet.

   a) Overall Species and Market Request

      (1) Row labels: NESPP4_GRP, SPECIES, MARKET

      (2) Values: Sum of # SAMPLES Requested, Sum of # LENGTHS Requested, Sum of # AGES Requested, Sum of # SAMPS Collected, Sum of # LENGTHS Collected, Sum of # AGES Collected, Sum of TOT WEIGHT Sampled, Sum of TOTAL LBS REPORTED, Sum of TOTAL # REPORTS, Sum of ALL TOTAL LBS REPORTED, Sum of ALL TOTAL # REPORTS

      (3) Copy and paste the results into the Overall Compare tab of the Request, Samples, Landings Compare spreadsheet.

   b) Full Request

      (1) Row Labels: NESPP4_GRP, SPECIES, MARKET, REGION, AREA, GEAR

      (2) Values: Sum of # SAMPLES Requested, Sum of # LENGTHS Requested, Sum of # AGES Requested, Sum of # SAMPS Collected, Sum of # LENGTHS Collected, Sum of # AGES Collected, Sum of TOT WEIGHT Sampled, Sum of TOTAL LBS REPORTED, Sum of TOTAL # REPORTS

      (3) Copy and paste the results into Full Request Compare tab of the Request vs. Samples –All Compare spreadsheet.

B. Calculate the percentages of requests filled

   1. Overall Comparison

      a) Calculate the percentage of requested lengths and ages collected by dividing the number collected by the number requested in the % LENGTHS Filled and % AGES Filled columns respectively.

(1) Filter for values of #DIV/0! and decide if they should be changed to N/A (no ages are requested) or if the row should be deleted (no request was made; exception: samples were collected).

(2) If the columns are not already formatted, format them so that they are percentages, values less than 50% are bold and red, and values greater than 100% are blue and bold.

b) Add a grand total row at the bottom of the spreadsheet.

c) Highlight rows with no possession species or rows where no request was made, but sample(s) were taken.

2. Full Request Comparison

a) Calculate the percentage of requested lengths and ages collected by dividing the number collected by the number requested in the % LENGTHS Filled and % AGES Filled columns respectively.

(1) Filter for values of #DIV/0! and decide if they should be changed to N/A (no ages are requested) or if the row should be deleted (no request was made; exception samples were collected).

(2) If the columns are not already formatted, format them so that they are percentages, values less than 50% are bold and red, and values greater than 100% are blue and bold.

b) Complete the LENS over/under column by subtracting the number of lengths requested from the number of lengths collected.

(1) If the column is not already formatted, format it so the any values less than 0 are bold and orange

c) Fill in the %SPP MKT Request Filled column with data from the corresponding species and market % LENGTHS Filled column in the Overall Compare tab.

d) Add a grand total row at the bottom of the spreadsheet.

e) Highlight rows with no possession species or rows where no request was made, but sample(s) were taken.

C. Identify Significant Landings

1. By Vessel

a) Using the Landings Data spreadsheet created above in A 3, create a pivot table to total the landings by PERMIT and DEALNUM

(1) Row labels: SPPNM, MKT_REQ, REGION, AREA_REQ, GEAR_REQ, PERMIT, DEALNUM, ST, PORTNM

(2) Values: Sum of LB, Sum of # RPTS

b) Copy and paste the values only of the results next to the table.

c) Add a LB/#RPTS column to calculate the average weight of landings per report by dividing the number of pounds reported by the number of reports

d) Sort the data by SPECIES, MKT_REQ, and LBS/# RPTS.

e) Use the average weight of a typical sample (see Appendix A for how to calculate) to highlight the rows as follows

(1) No area, &/or LBS/ # RPTS less than estimated sample weight

(2) 1-5 reports and LBS/ # RPTS approximately equal to or greater than estimated sample weight

(3) 6-29 reports & LBs/ # RPTS approximately equal to or greater than estimated sample weight

(4) 30 or more reports & LBS/ # RPTS approximately equal to or greater than estimated sample weight

f) Copy and paste the data into the Significant Landings tab of the Request vs. Samples –All Compare spreadsheet and delete the rows that fall into category 1 from 1 e above.

2. By Dealer

a) Using the significant landings by vessel data, create a pivot table next to the data with the below criteria.

(1) Row labels: SPPNM, MKT_REQ, REGION, AREA_REQ, GEAR_REQ, DEALNUM, ST, PORT

(2) Values: Sum of LB, Sum of # RPTS

b) Copy and paste values only of the pivot table results next to pivot table and delete the original pivot table.

c) Highlight rows according to below criteria

(1) 1-5 reports

(2) 6-29 reports

(3) 30 or more reports

D. VTR Landings and Associated Graphs

1. Use the VTR Landings SQL script to compile VTR landings data for the specified time period.

a) Specify the time period and the VTR species codes for the requested species in the upper section of the script. Create the table and run the section of the script that assigns the appropriate SPECIES, STOCK AREA, GEAR, and REGION information to the data.

b) Copy and pasted the results into the VTR Landings tab of the Request vs. Samples –All compare spreadsheet.

2. Use the Clam VTR Landings SQL script to compile the surf clam and ocean quahog landings data.

a) Specify the time period in the upper section of the script, create the table, and run the section of the script that assigns the appropriate SPECIES, STOCK AREA, GEAR, and REGION information to the data.

b) Copy and paste the results into the VTR Landings tab of the Request vs. Samples –All compare spreadsheet.

3. Used the compiled data to create the following pivot charts

a) Landings by species and gear

(1) Chart specifications

(a) Axis fields: SPPNM

(b) Legend Fields: GEAR GRP

(c) Values: sum of LB

(2) Change chart type to 3-D Columns

(3) Add Chart titles, axis titles, and other necessary text boxes

b) Landings by species, stock area, and gear

(1) Chart specifications

(a) Axis fields: SPPNM, STOCK ABBR

(b) Legend Fields: GEAR GRP

(c) Values: sum of LB

(2) Change the chart type to 3-D Columns.

(3) Add Chart titles, axis titles, and other necessary text boxes.

E. Post the finished spreadsheet to the Sampling Request- Samples- Landings Compare page of the NE Port Biological Sampling Program Wiki page (*https://www.st.nmfs.noaa.gov/confluence/pages/viewpageattachments.action?pageId=42501529&metadataLink=true*).

III. Sampling Review Progress Report

A. Create summaries of sampling related JIRAs.

1. NE Port Biological Sampling Program (NESPBSP)

a) Use the Issue Statistics JIRA gadget and the 2012 NEPBSP and the 2013 NESPBSP filters (Appendix B) to update the totals in the top table in the NEPBSP JIRAs tab of the Sampling Program Review Progress Report spreadsheet.

b) Update the number of Currently Open/ In Progress JIRAs at the top of the tab.

c) Use the Created vs. Resolved Chart JIRA gadget and the 2013 NESPBSP filter to update the 2013 JIRAs table.

(1) Maximize the gadget and edit the table if necessary to include all days of the current year.

(2) Add the necessary week(s) and the corresponding numbers of created and resolved JIRAs to the 2013 JIRAs table.

(3) Verify that the totals match the numbers in the top JIRA table.

d) Copy and paste the results of the Two Dimensional Filter Statistics: 2013 NEPBSP (+ active 2012) (Appendix B) JIRA gadget into the 2013 JIRA Status table.

2. NER Compliance (NERC)

a) Update the number of Currently Open/ In Progress JIRAs at the top of the tab using the Issues Statistics JIRA gadget and the Samp NERC filter (Appendix B).

b) Add the necessary week(s) and the corresponding numbers of Moved and Resolved JIRAs to the ALL NERC table.

c) Copy and paste the results of the Two Dimensional Filter Statistics: Samp NERC JIRA gadget into the 2013 JIRA Status table.

B. BSMDBS Review

1. Update the time period at the top of the BSMDBS Audits tab of the Sampling Review Program Progress Report spreadsheet.

2. Copy and paste the summary table of potential issues broken down by staff type from the most recent BSMDBS Audit Spreadsheet.

3. Create a summary table of potential issues without staff type.

C. Samples Requested vs. Collected

1. Each sampling quarter will have its own tab in the Sampling Review Program Progress Report spreadsheet.

a) Sampling quarters in progress: Copy and paste the number of LENGTHS REQUESTED, AGES REQUESTED, LENGTHS COLLECTED, AGES COLLECTED, % LENGTHS Filled and % AGES Filled from the Overall Compare tab of the most recent Request, Samples, Landings Compare spreadsheet into a tab for that quarter.

b) Completed sampling quarters: If not already done, update completed quarters' tabs using the final Request, Samples, Landings Compare spreadsheet for that quarter.

D. Post the finished spreadsheet to the Program Progress Reports page of the NE Port Biological Sampling Program Wiki page (*https://www.st.nmfs.noaa.gov/confluence/pages/viewpageattachments.action?pageId=42501523&metadataLink=true*).

Appendix A: Calculating Average Weight of Samples

1) Use the Ave WT SQL script from the NE Port Biological Sampling Program wiki page (https://www.st.nmfs.noaa.gov/confluence/display/PBSP/NE+Port+Biological+Sampling+Program) under Program Management/Quality management/Scripts.at script) to pull 2009 and later sampling data for samples with the required amount of lengths +/-10 and sample weight >0.
   a) If no samples meet these criteria for a given species and market category, use data from previous years.
   b) If no samples have the required number of lengths +/-10, use the existing data to estimate a per fish weight and extrapolate the weight of a full sample.
2) Species not requiring a weight
   a) Scallops - use 500 LBS
   b) Skates- use 300 LBS
   c) Surf clams and Ocean Quahogs- no landings should be excluded based on weight since the landings numbers are in bushels.

Appendix B: JIRA Filters

2012 NEPBSP:

project = "NE Port Biological Sampling Program" AND type != Task AND created < "2013/01/01"

2013 NEPBSP:

project = NEPBSP AND created > "2013/01/01" AND type != task

2013 NEPBSP (+active 2012):

project = NEPBSP AND (created > "2013/01/01" OR status in (Open, "In Progress", Reopened)) AND type != task

Samp NERC*:

project = "NER Compliance" AND created >= "2012/11/08" AND (reporter = "Aaron.Dieckerhoff" OR reporter = "alissa.wilson" OR reporter = "jodie.york" OR reporter = "joshua.o'connor@noaa.gov" OR reporter = "katie.almeida" OR reporter = "Marc.Renaghan" OR reporter = "sara.petrochic@noaa.gov" or reporter = "Robert.Bollmann") AND key >= NERC-309

* Change or add reporter names as necessary