Speaker's Notes

Slide 1 --- None

Slide 2 --- None

Slide 3

In this presentation, we will provide

- The theoretical underpinnings of MiST and review the methodology employed
- Describe the assumptions built into MiST for analyses of PCB Mixtures
- Describe the key information users will need to analyze similarity of PCB mixtures
- Illustrate the user inputs required and the outputs

Slide 4

- MiST is tool to analyze similarity among PCB mixtures that was developed in support of EPA's IRIS PCB assessment of noncancer human health hazards
- The goal was to develop a mixtures similarity tool for risk assessment scientists that was accessible and easy to use
- Therefore, this mixtures similarity tool was developed in Microsoft Excel with the objective of providing an intuitive and easy to use user interface to aid risk assessors in determining if "candidate mixtures" and "reference mixtures" are similar from a toxicity standpoint.

Slide 5

- Here is an illustration of the key question MiST is designed to answer
- On the left are two reference mixtures, A and B. These are defined as mixtures of known congener composition for which BMDs are known. For the purpose of this illustration, each reference mixture harbors only 4 of the 209 PCB congeners
- Reference mixtures A and B harbor overlapping but non-identical congeners.
- Shown on the right is a candidate mixture of interest. Again, for clarity, the mixture is shown with only 3 congeners that partially overlaps congeners in both reference mixtures.
- MiST is designed to answer a fundamental question: Is the candidate mixture sufficiently similar to the reference mixtures such that the reference could be used as a toxicity surrogate?

Slide 6

- Mist defines reference mixtures as those for which estimated effect levels (e.g., benchmark doses (BMDs)), along with variance information for these estimates that can be or have been derived. Reference mixtures could be industrial or environmental mixtures that have known BMDs for one or more health effects.
- The candidate mixture is the mixture of interest for risk assessment. For example, environmental mixtures or mixtures characterized in biological samples. This mixture might lack adequate dose-response data for deriving estimated effect levels.
- MiST will determine if the candidate is sufficiently similar to one or more reference mixtures.

Slide 7

- More specifically, MiST can 2 questions about the candidate mixture:
- First, is a given reference mixture "sufficiently similar" to the candidate mixture such that the reference mixture could be used as a toxicological surrogate?
- Second, if more than one reference mixture is "sufficiently similar" to the candidate mixture, which reference mixture is the most appropriate toxicological surrogate?

Slide 8

- MiST methodology was adapted from an approach to determine sufficient similarity for mixtures by Marshall et al., developed by a group led by Chris Gennings.
    - Two major modifications were introduced for MiST:
    - Whereas the Marshall approach used Taylor series expansions to evaluate the distribution of distances between reference and candidate mixtures, MiST uses the Monte-Carlo method. The advantage is that the Monte Carlo approach does not rely on the BMD estimates being normally distributed.
- To improve its utility for the IRIS PCB assessment, Mist was amended such that users can compare one candidate mixture to either one reference mixture or multiple reference mixture(s).

Slide 9

- MiST evaluates similarity between mixtures by calculating the average Euclidean distance between reference and candidate mixtures.
- In MiST, this critical values is typically set to the absolute value of the difference between the effective dose (ED) and median BMD-value.
    - If the 95th percentile of the "distance" is less than the critical value, then the reference is "sufficiently similar" to the candidate
    - If several reference mixtures are compared to the same candidate mixture, the reference mixture with the smallest mean distance is distinguished as the best match to the candidate mixture

Slide 10

- This is a visual depiction of the MiST approach
- This illustration depicts a two-dimensional representation of a 209-dimensional calculation that MiST executes for each congener
- In this illustration, the reference and candidate mixtures are shown as the different hashed red, blue and green lines, and length of each dashed line reflects the uncertainty of the BMD.
- Each purple and black line represent the result of a single Monte Carlo iteration, which are ranked by length.
- Mixtures are considered similar if the distances between the candidate and references are less than the critical value for at least 95% of Monte Carlo iterations

Slide 11

- This is a visual depiction of the MiST approach
- This illustration depicts a two-dimensional representation of a 209-dimensional calculation that MiST executes for each congener

- In this illustration, the reference and candidate mixtures are shown as the different hashed red, blue and green lines, and length of each dashed line reflects the uncertainty of the BMD.
- Each purple and black line represent the result of a single Monte Carlo iteration, which are ranked by length.
- Mixtures are considered similar if the distances between the candidate and references are less than the critical value for at least 95% of Monte Carlo iterations

Slide 12

- This is a visual depiction of the MiST approach
- This illustration depicts a two-dimensional representation of a 209-dimensional calculation that MiST executes for each congener
- In this illustration, the reference and candidate mixtures are shown as the different hashed red, blue and green lines, and length of each dashed line reflects the uncertainty of the BMD.
- Each purple and black line represent the result of a single Monte Carlo iteration
- Mixtures are considered similar if the distances between the candidate and references are less than the critical value for at least 95% of Monte Carlo iterations

Slide 13

MiST was developed specifically to analyze 209 components in order to apply this to PCB mixtures. Thus, certain assumption have been built into the tool

- First, The chemical mixtures consist of 209 congeners. Inherit in this assumptions is that there are no other chemicals in the mixture and that the congener mass fractions sum to one
- Second, the health effects for all reference mixtures and any candidate mixtures in a single analysis are comparable. It is the responsibility of the user to assure that each BMD refers to the same health effect.
- Third, the fundamental principle used to establish sufficient similarity is based on the difference between the candidate and reference BMDs (calculated as the weighted Euclidean distance between them)

Slide 14

1. Several additional assumptions are built into MiST regarding selection of BMDs for health effects under study.
2. First, each reference mixture has a BMD, or BMD-like estimate, with an associated uncertainty distribution, for some health effect
    - The BMD of a mixture is defined as the weighted sum of the BMD of its congeners
    - MiST also assumes there is no synergism or antagonism among congeners. Stated another MiST assumes there is no dose additivity
3. Second, MiST assumes the BMD distribution for the candidate mixture is independent of those for the reference mixtures, meaning the covariance between the two distributions is assumed to be zero

Slide 15

Finally, built into MiST are certain assumptions regarding the congener characteristics of PCB mixtures and how they are specified in the tool.

- First, the congener mass fractions sum to 1 and are specified without any estimate of their potential uncertainty or error
- Second, the relative potency of each congener is specified without an estimate of uncertainty or error
    - Of course, there may be some level of uncertainty for both mass fractions and relative potencies depending on the data available for each mixture under study. However, at this time MiST does not include the functionality required to incorporate these uncertainties into the similarity analysis.
- Finally, MiST assumes the "relative toxicological potencies" of all congeners in a mixture should sum to 209. If potencies do not sum to 209, MiST will autoscale potencies such that they do sum to 209

Slide 16

The amended method employed by MiST to determine mixtures similarity requires, at a minimum, 3 inputs from the user:

- Mass fraction of each congener in the candidate and reference mixtures
- BMD ± SD or BMD Cumulative Distribution Function (CDF) for each reference mixture
- Effective dose (ED) for each reference mixture

Note that while congener toxicological relative potencies can be added if known, these data are not required to run an analysis.

Slide 17

In this next section, we will provide an overview of the information users need to use MiST.

- MiST can answer the question of mixture similarity for several data availability scenarios.
    - Are relative potencies available for the congener constituents of the mixture?
    - Does the candidate mixture have BMD information? The answer to this question determines whether the analysis will be a data rich or data poor scenario.
- The 2 questions shown here are designed to aid the user in assessing their data availability scenario:

Slide 18

- This illustration depicts the 6 possible data availability scenarios that can be used in MiST based on the 2 questions posed in the previous slide
- It is designed to aid the user in determining evaluating whether the analysis represents a data rich or a data poor scenario, and the level of confidence in data available for a MiST analysis.
- The first question is whether relative potencies are available for all the mixtures of interest.
    - For example, if the answer is YES, it is a weighted analysis and we follow top section of the tree for the available scenarios.
    - If the answer is NO, MiST can still evaluate similarity by assuming equal potencies among congeners and we follow the bottom section of the tree
- The second question relates to information available for the candidate mixture.
    - If BMD is not available, the analysis is considered data poor.
    - If the BMD is available, the analysis is considered data rich.

- The level of confidence is indicated by the color intensity of the boxes and arrows on the left-hand side of the visual, with more intense colors reflecting greater confidence.
    - Overall, confidence is highest for data availability scenarios at the top of the decision tree and decreases as you move to scenarios lower in the tree.
    - The highest confidence analysis will be determined by whether the user can run a weighted analysis, and whether the user has BMD information for the candidate mixture.
    - Confidence is also increased by the information available for the BMD variance

Additional Note

– Marshall et al., was originally designed for the data poor scenario

Slide 19

- MiST is an excel-based tool and data inputs will be added into forms set up in the excel workbook
- Before using MiST, the user should gather required information, and if available, optional information:

Required information:

- Mass fraction of each congener in the candidate and reference mixtures
- BMD ± SD or BMD Cumulative Distribution Function (CDF) for the reference mixture
- Effective dose for the reference mixture. : The difference in estimated BMD and effective dose (ED) for reference mixture is used to calculate the critical value (Δ)
    - There is an option to enter a user-specified critical value. However, this is not recommended

Optional Data:

- BMD for the candidate mixture (which will determine if the analysis will run as a data rich or data poor scenario)
- The relative toxicity of different constituents in the mixture. If this information is not available, MiST assumes equal potencies for all congeners, indicating a data poor scenario

Slide 20

- For the demonstration, we will use the example discussed in the earlier talk by Laura Carlson and Jeff Gift at EPA
- BMD modeling on a neurological endpoint (landing foot splay) from the paper by Freeman et al that evaluated neurotoxicity of aroclor mixtures
- Endpoint was evaluated after 28 weeks of dosing
- Calculated BMD with CDF
- Used Neurotoxicity TEQs as congener toxicological potency values
- Mass fraction data for the aroclors derived from the ATSDR tox profile for PCBs
- Assessed similarity between a candidate mixture (Aroclor 1254) and three reference mixtures (Aroclors 1016, 1242, and 1260)

Slide 21

- Upon opening the tool for the first time, a yellow banner will appear that asks the user to enable macros.
- This may appear as a modal dialog box if using MiST on an apple computer
- Users must click to enable macros

- Any other excel books with macros must be closed before using MiST
- For optimal performance, close other excel workbooks while using MiST
- Animation: The version number and release date are shown in the blue header

Slide 22

- Before starting the demo, we would like to present a short overview of each tab
- The background tab provides a very short overview of MiST methodology
- Specifically, the methodology employed to answer two questions related to similarity and ranking of reference mixtures relative to the candidate mixture
- The information on the background tab is a summary of the methodology, and users should consult the User Guide for full methodological details, principles, and validation tests

Slide 23

- The quick start guide provides a condensed set of instructions for conducting a mixtures similarity analysis
- It is intended as a quick reference guide, and users should consult the MiST User Guide for details on using the tool

Slide 24

- Most of the data will be entered into the data repository tab
- Here, the user will specify the number of mixtures, the mixture names, the BMD type, (mean and SD or CDF values), and the ED
- Also, mixture composition data (congener mass fractions) are also entered on this tab

Slide 25

A note on entering CDF information

- If CDF values are entered into the clipboard, they can be automatically copied into the Data Repository Tab using the “Copy CDF” button
- The BMD values should be entered as a whole number or decimal greater than zero in non-decreasing order.
- If a user does not have BMD values for every percentile, a piecewise linear interpolation function can be performed to estimate the values for the missing percentiles of the CDF using “Fill” button

Slide 26

If not all CDF values are known, the user may use the fill button to interpolate missing values
There are some assumptions implicit in the interpolating function in MiST

1. All values must be specified
2. BMD values can only be entered for the whole number percentiles as decimals or whole numbers.
3. No value may be negative
4. No value may be smaller than a value at a lower percentile
5. If two or more values are entered, the user may use the “Fill” button to populate the rest
    - Each integer percentile gap (between two specified values) is filled by linear interpolation

- Interpolation can be used to fill towards the minimum value, the maximum value, or in both directions
- Values above the highest specified percentile are filled using the same slope (increase per percentile) as occurs between the two largest specified values. The same logic applies to interpolation to the minimum

6. Users can also easily clear data using the “Erase” button

Slide 27

- The settings tab is where users enter key information about the run
    - Whether congener potencies are known
    - Whether the analysis is data rich or data poor
    - Here the user will also specify which mixture is the candidate, and which mixture or mixtures are designated as reference mixtures
    - Finally, if congener potency information is available, it should be entered here

Slide 28

The results tab provides basic information about the candidate mixture, and ranks the similarity of reference mixtures relative to the candidate mixture

Slide 29

There are two additional tabs in the tool: The Clipboard tab and the Example Data tab:

- User may enter CDF data in the Clipboard tab
- The clipboard enables user to copy or enter data from different various sources (e.g., EPA BMDS software output, data from published manuscripts, other spreadsheets etc.)
- Provides a space for users to review and perform QC on CDF values
- Entering data into the clipboard first ensures all values are in an acceptable format for pasting into the Data Repository tab

Slide 30

- So now let’s start the demonstration.
- To begin, first navigate to the Data Repository Tab
- Click the “New Mixture button” to add additional mixtures
- Here, we will the candidate mixture as Mixture #1, and the reference mixtures as Mixtures 2, 3 and 4
- You can use the “Reorder” button to change the order of the mixtures listed in the tab

- First, enter the mixture name for the first mixture
- Enter the ED value
- Then, select the BMD type using the dropdown menu

Slide 31

- After choosing the BMD type, MiST will automatically move the cursor to the cells below the mass fraction table
- Enter the appropriate values for BMD
- If CDF is selected, a new table will appear to enter CDF values

- If CDF values were copied to the clipboard, use the “Copy CDF” button to automatically copy values into the new table

Slide 32

- Next, the user should enter the congener mass fraction data
- Congener mass fractions are defined as mass of the congener divided by total mass of all congeners in the mixture - $W_i = (m_i/m_T)$
- MiST automatically calculates W_i given user-inputted m_i values
- Keep in mind that a value should be entered for each congener (you will receive an error if there are blank values for any congener)

Slide 33

- The last step is to lock the mixture
- The analysis will not run until all mixtures are locked
- It should be evident when the mixtures are locked because the cells now appear with a white background. Tip - The blue cells indicate those that can be edited. After locking, the user cannot edit values without first unlocking the mixture.
- Repeat steps 1-6 for each mixture

Slide 34

- Once all data is entered on the data repository tab, the user should navigate to the Settings tab
- Here the user will define the type of analysis that will be conducted
    - This includes indicating whether congener potencies are known
    - If they are unknown MiST assumes equal potencies
    - Here the user also specifies whether a data rich or a data poor analysis will be conducted
    - Next, we specify which mixture is the candidate, and the number and names of one or more reference mixtures
    - Finally, we enter the congener potencies, if known. If potencies are unknown, MiST assumes equal potencies for all congeners
    - If congener potencies are entered, the column for equal potencies is blacked out

Slide 35

- Finally, when all the data is correctly entered, the analysis will begin once the “Run Analysis” button
- Once the analysis is run, a dialog box will appear indicating the analysis is complete
- When the user clicks “OK” MiST will automatically navigate to the results tab
- Note this tab has additional functionalities. It allow users to edit their run, save their run, or clear data for a new run. We suggest that users save their runs under a new filename, to avoid overwriting the original file.

Slide 36

- Run settings are shown on the left
- After similarity ranking, the mixture nearest to the candidate is shown along with the candidate BMD and ED values

- If the mixture name and distance are in green, this indicates MiST determined the nearest mixture is toxicologically similar
- If the name and distance are in red, MiST determined the mixture is not sufficiently similar to the candidate to ascribe toxicological similarity for the given health effect

Slide 37

- For the reference mixture (s), MiST returns the following information:
    - BMD - If mean and SD were entered, the mean BMD is returned. If a CDF was entered, the 50^{th} percentile is returned.
    - ED: the ED associated with a biological effect of the indicated reference mixture is returned.
    - Delta: the critical value for the analysis, set to |ED - BMD| by default.
    - Dw (Mean): the average Euclidean distance between the reference mixture and the candidate mixture.
    - Dw (Upper 95th): the 95th percentile of the Euclidean distance between the reference mixture and the candidate mixture.

Slide 38

How to Interpret MiST conclusions
Acceptable - the Dw (Upper 95th) value for the reference mixture is less than the critical value (Delta); consider the reference mixture as SIMILAR to the candidate mixture.

Not Acceptable - The Dw (Upper 95th) value for the reference mixture is greater than the critical value (Delta); in this case, we are unable to claim that the reference mixture is similar to the candidate mixture.

Rank: when several reference mixtures are considered similar to the candidate mixture, the rank indicates which reference mixture is the best match to the candidate mixture. This is determined by comparing the Dw (Mean) for each reference mixture

Note: Because Monte Carlo simulations are used to establish the mean Dw, there is some intrinsic variability between different runs of the same inputs.

Slide 39

To summarize,
MiST is a Microsoft Excel-Based tool developed to support risk assessment of chemical mixtures relevant to human health. It is designed to be user friendly and accessible to a wide audience

In order to assess similarity, MiST uses a Monte Carlo-based approach to account for the uncertainty in candidate and reference BMDs

MiST ascertains the degree to which “candidate mixtures” and “reference mixtures” are similar from a toxicity standpoint.. To do this, MiST measures the distance between the candidate and reference mixtures and compares it to a critical value representing a user-defined similarity boundary.

Slide 40