POLICY-RELEVANT LESSONS LEARNED FROM PHASE I/II AIR TOXICS ANALYSES

WHITE PAPER
STI-903553-2479-WP

By:
Hilary R. Hafner
Michael C. McCarthy
Steven G. Brown
Paul T. Roberts
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
Lake Michigan Air Directors Consortium
2250 East Devon Avenue, #250
Des Plaines, IL 60018-4511

January 26, 2004

POLICY-RELEVANT LESSONS LEARNED
FROM PHASE I/II AIR TOXICS ANALYSES

Two data analysis efforts were funded in 2001-2003 to guide the design of a national air toxics monitoring program (Battelle Memorial Institute and Sonoma Technology, 2003; Bortnick et al., 2001; Bortnick et al., 2003). While these analysis efforts used historical (Phase I) and pilot city (Phase II) air toxics databases to answer a series of questions related to network design, several policy-relevant questions were investigated as well. The following discussion summarizes the findings from the 2001-2003 and related projects in the context of policy questions developed for the third phase of the analysis of air toxics data (Hafner et al., 2004). The questions are grouped into broad categories; some questions were not addressed in the previous work and are identified as such here. The analyses planned in 2004 are designed to further address this list of questions.

How do we identify whether a community's data can be used to address policy-related questions?

Can policy questions be addressed with the available and planned data?

- Contributions from likely source types were identified with source apportionment techniques using the Detroit pilot city data (e.g., use of 1-in-6-day samples, identification of more than 18 air toxics); these analyses are important in understanding what control strategies should be applied and assessing their potential effectiveness.

- Benzene data from the historical archive have been used to illustrate the quantitative decreases in ambient benzene data due to the introduction of reformulated gasoline (Main et al., 1998; Main, 2002).

- Long-term (i.e., more than ten years) sampling provides data for trend analyses so that changes in ambient concentrations can be investigated with respect to control measure changes. Investigation of annual trends in concentrations was not specifically performed with the air toxics archive data but is planned by Hafner et al. (2004).

What is our confidence in the data?

- Detailed quality assurance information is missing from the historical database, such as precision and accuracy of both the sampling and the laboratory analyses, interlaboratory comparison results (i.e., analytical precision), and audit results. The lack of these complementary data makes it difficult to estimate the uncertainty in the air toxics concentrations.

- Confidence in air toxics data is a function of the pollutant, sampling and analysis methods, and analysis objectives. For example, initial exploration into the implications of sampling and analysis method differences to data quality and data comparability showed that, in the absence of method precision and accuracy information, a high ratio of median concentration to median minimum detection limit (MDL) is an indication of good data quality. Several compounds had medians more than ten times the MDL (indicating

1

high confidence in the data): formaldehyde, PM_{10} lead, PM_{10} manganese, lead total suspended particulate (TSP), benzene, carbon tetrachloride, and methylene chloride. In contrast, the pilot city study indicated serious data quality questions regarding acrolein – "the Pilot Study confirmed that we're not positioned to provide routine acrolein data of sufficient quality" (Battelle Memorial Institute and Sonoma Technology, 2003); these results indicate that confidence in the historical record for acrolein is also questionable.

- Significant data quality issues with the historical archive were addressed by Battelle and included extensive efforts to organize, flag, and more fully document the data. Examples of Level 1 data validation actions included

  – identifying data that lacked method information (such as unknown MDL, unknown sampling artifacts/precision) that may impact evaluation of the historical trends;

  – addressing ambiguous or missing information (e.g., missing units, unclear units such as ppb vs. ppbC that may render the data useless); and

  – noting missing supporting information (i.e., metadata) including undocumented or inconsistent sampling objectives or lack of specific site descriptions (typically just a number referring to a broad category such as urban or rural) that may impair inter-site comparisons or other uses of the data.

  These data validation actions are vital to provide a more robust database for future analyses.

- Additional analyses will be carried out in 2004 to characterize confidence in the historical data.

Which toxics species are adequately represented in the database (e.g., national/regional or urban/rural coverage)?

- Focusing on the pilot study's 18 core air toxics, more than 60% of samples for benzene ($>$ 300,000 samples), 1,3-butadiene ($>$ 88,000 samples), formaldehyde ($>$ 44,000 samples), and lead TSP ($>$ 143,000 samples) are valid and have date, site, method, and MDL documentation. Other core air toxics with more than 10,000 24-hr average valid observations include acetaldehyde, arsenic ($PM_{2.5}$, PM_{10}, TSP), beryllium TSP, cadmium (PM_{10}, TSP), chromium ($PM_{2.5}$, PM_{10}, TSP), lead ($PM_{2.5}$, PM_{10}), manganese ($PM_{2.5}$, PM_{10}, TSP), nickel ($PM_{2.5}$, PM_{10}, TSP), 1,2-dichloropropane, carbon tetrachloride, chloroform, methylene chloride, tetrachloroethylene, trichloroethylene, and vinyl chloride. Many data values for some species (i.e., vinyl chloride, acrolein, 1,3-dichloropropane, chromium VI, trichloroethylene, and chloroform) are below the listed MDL.

- The historical air toxics database does not contain available data that are useful to the investigation of air toxics, such as the majority of Photochemical Assessment Monitoring Stations (PAMS) data sets (which contain benzene, and sometimes 1,3-butadiene, formaldehyde, and acetaldehyde) and speciated PM data sets (which contain chromium, lead, nickel, manganese, EC and mass). Therefore, analyses of toxics data need to consider other data sources.

- Data for species that are not air toxics (e.g., organic carbon [OC], elemental carbon [EC], sulfate, nitrate) would be useful in data validation (e.g., following procedures outlined in the PAMS and $PM_{2.5}$ data analysis workbooks [Main and Roberts, 2000, 2001]) and data analysis (e.g., source apportionment) efforts. These data should be added to the air toxics database.

How should missing data and data below detection levels be treated, and how do different data treatments affect data analysis and modeling results?

- The selection of the treatment method for data below the MDL can have a significant impact on analysis results if a large amount of data is below the MDL. In the computation of an annual average (AA),
  - ignoring data below the MDL results in an AA that is biased high;
  - replacing the data below the MDL with the MDL value results in an AA that is biased high; or
  - replacing the data below the MDL with zero results in an AA that is biased low.

- The process of using reported values below the MDL (when available) and using MDL/2 for nondetect data provides a defensible annual average for data sets with up to 50% of data below the MDL. If more than 50% of the data are below the MDL, the annual average can be biased by the choice of MDL substitution.

- Analyses indicated that values below the MDL should be reported to (flagged, but not censored from) the database, and used.

- Analyses will be performed in 2004 to further investigate MDL substitution schemes and their effect on annual averages. For the 18 core air toxics, cancer benchmark values will be compared to representative MDL values. For those species with representative MDLs equal to or greater than the benchmark values, new analytical techniques will be needed to adequately assess possible health risks. For those species with representative MDLs sufficiently below the benchmark concentrations, treatment of data below the MDL will not be a critical issue for health assessments.

Are data useful even if they do not meet QA/QC and data completeness criteria?

- The importance of data completeness is a function of the variability of pollutant concentrations. For example, fewer samples may be needed to compute an AA for a toxic air pollutant with a long atmospheric lifetime (and low current emissions rates) such as carbon tetrachloride because this pollutant showed little spatial and temporal variability. Data completeness is more important for the more highly variable air toxics.

- Data that do not meet QA/QC criteria are not useful for quantitative analyses.

How good are the historical and pilot city data and what are the criteria for determining "good"?

- The historical and pilot city data were used for a wide range of analyses to explore sources of pollutant variability and ultimately to establish guidelines for the development of a national toxics network. The data were of sufficient quality and completeness to be

3

useful for this purpose. However, the data for some HAPs may not be of sufficient quality and completeness for statistically demanding analyses such as trend analyses.

- The criteria for determining "good" will have to be developed based on the data analysis objectives. For example, is data precision good enough for analysts to observe a 15% change in concentration between two consecutive three-year periods (i.e., one of the data quality objectives for the national air toxics trends sites [NATTS] program)? The pilot city showed that acetaldehyde, formaldehyde, benzene, arsenic TSP, and chromium TSP had coefficients of variation (CVs—a measure of analytical precision) of 10% or less, and thus would be sufficient for this trend analysis. However, other toxics had higher CVs.

- The pilot city data showed that high MDLs (with respect to cancer benchmark concentrations) and inter-laboratory and inter-study comparability issues for some species (such as hexavalent chromium) need to be addressed in order for the data to be useful in future analyses.

What are the air toxics concentration levels nationally and locally?

Based on case studies, what can we say about air toxics concentrations on the urban scale?

- Analyses showed that where air toxics concentrations are not dominated by local emissions, urban concentrations can be characterized by few (or one) sampling sites within a city. For some species, spatial variability was high; thus, one site would be inadequate.

- In most of the case studies using the archive data, high spatial variability in a county was caused by local point sources (and source-oriented monitoring).

- Additional investigations will be made in 2004 with the historical and pilot city data sets to further understand urban air toxics concentrations. For example, we will categorize the species based on photochemical lifetimes to help future data analysts determine which species require multiple monitoring sites.

How can concentrations from a small network be extrapolated to other areas? and How representative are the existing monitoring sites (i.e., how broadly [spatially] can the annual average concentration at a particular site be applied)?

- Spatial representativeness depends on pollutant and location—the concentrations of some species with long lifetimes, such as carbon tetrachloride, will be easier to extrapolate from one area to another because emissions are low in the United States and both spatial and temporal variability in this pollutant's concentrations were low compared to other air toxics.

- Additional spatial analyses will be performed in 2004 to further understand representativeness.

4

Is more detailed "standard metadata" needed to better define the specific, micro-scale characteristics of air toxics monitoring sites?

- According to the analyses, more detailed information is desirable. For example, extensive investigation was required to track down the details of site characteristics and monitoring objectives to help explain data anomalies; these details were not readily available in the data set. These anomalies need to be addressed in both data validation and data analysis in order to understand whether the anomalies are due to local sources or atmospheric variability of regional sources. Thus, metadata are vital.

- More than just the toxic species are needed in validation and analysis. Air quality measurements of pollutants not listed as HAPs, such as OC, EC, sulfate, and nitrate, would allow data analysts to check mass and ion balances for PM data. Other VOC species (such as the PAMS target analytes) are useful for the investigation of the validity of selected HAPs (e.g., comparing benzene to acetylene). For source apportionment, species such as iron, calcium, OC, EC, hydrocarbons, etc. improve the identification of emission sources.

What is meant by "background" levels?

- Background concentrations can be defined as global, regional, and/or operational:

  – The definition of global background is the mean concentration of a compound measured at a remote site not impacted by local or regional emission sources. Example global remote sites include islands in the middle of the ocean or polar regions.

  – The definition of regional background is the mean concentration of a compound measured at a site not impacted by local sources. A regional background site might be one located in a coastal area with predominantly onshore flow, or a remote continental site uninfluenced by local sources.

  – The definition of operational background is the mean concentration of a compound measured at an upwind site. The concentration of reactive compounds may need to be adjusted downward to account for the residence time until collection at the downwind site. This operational definition may include urban or rural sites upwind of other monitoring stations, since that air is the "background" for sites downwind.

  – In the Assessment System for Population Exposure Nationwide (ASPEN) model, background concentrations are defined to account for natural sources, sources farther than 50 km from the monitoring station, and unidentified sources. This is close to the operational definition defined above, although it also includes unidentified sources and local natural sources. In our planned analyses, we would rather not use this definition because it involves adjusting operational background concentrations to account for deficiencies in emission inventories which may be independently corrected.

- Thus, the global definition identifies the lowest possible concentrations occurring in the atmosphere, whereas the operational definition may be much higher depending on proximity to upwind sources and the nature of the pollutant. The operational definition of background concentration implies that backgrounds vary significantly by site location.

- Background concentrations will be investigated further in 2004.

How can background levels be estimated?

- Background levels can be estimated in a number of ways, depending on pollutant and definition used:

  - Global background concentrations of many of the 18 target species are routinely measured at remote monitoring stations operated by National Oceanic and Atmospheric Administration (NOAA) Climate Monitoring and Diagnostics Laboratory (CMDL).

  - Regional background concentrations may be estimated from rural monitoring stations for compounds with relatively long atmospheric residence times. The concentrations of these compounds should exhibit high spatial correlations (i.e., low spatial variability).

  - Operational background concentrations consist of the global background concentrations plus regional, natural, and upwind sources of pollutants. Use of upwind monitoring stations, chemical lifetimes, and ambient monitoring data can be used to help identify operational background concentrations on a site-by-site basis. This approach may not be practical for all monitoring sites in the historical database.

- Background concentrations will be investigated further in 2004.

What are reasonable estimates of background levels?

- The 2004 analyses will determine reasonable estimates of global backgrounds from the scientific literature and from remote/rural monitoring sites. Reasonable estimates of operational backgrounds will be calculated using known global backgrounds and available pollutant- and location-specific information.

What does a broad national assessment say about air toxics concentrations across the country?
How do urban area concentrations compare with those in nearby rural areas?
Are there typical urban and rural profiles (fingerprints)?
How does exposure to mobile-source toxics vary as a function of distance from roadways?

- These questions were not specifically addressed in 2001-2003 but we will attempt to address them in 2004.

How do air toxics concentrations vary temporally?

What can we say about the variation in air toxics concentrations on a yearly (trends), seasonal (or monthly), day-of-week, and hour-of-day basis?

- Significant seasonality was observed for formaldehyde and acetaldehyde at more than half the sites; these had higher concentrations in the summer, likely due to the photochemical production of these pollutants from other VOCs. Benzene concentrations were higher in the fall-winter period, likely due to lower mixing heights and less photochemical destruction than in the summer. Other HAPs typically did not exhibit statistically significant seasonal patterns, or the patterns were site-specific.

- By day of week, no pattern was observed for acetaldehyde, trichloroethylene, tetrachloroethylene, formaldehyde, chloroform, benzene, methylene chloride, carbon tetrachloride, and 1,3-butadiene concentrations. However, other studies have shown day-of-week trends for urban morning benzene concentrations (i.e., lower concentrations on Sundays compared to other days).

- Diurnal variability was low for acetaldehyde, benzene at downwind/upwind sites (non-urban), methylene chloride, chloroform, 1,3-butadiene, carbon tetrachloride, trichloroethylene, and tetrachloroethylene. In contrast, formaldehyde and urban benzene had significant diurnal variability.

- Supplemental data could also be useful in the investigation of trends. For example, speciated $PM_{2.5}$ data from the IMPROVE network can be used to determine seasonal and interannual trends in As, Cr, Mn, Ni, and Pb at rural sites.

- Investigation of annual trends in the concentrations was not specifically performed with the air toxics archive data but is planned for 2004.

How do we characterize "spikes" in ambient concentrations?

- High concentration "spikes" were identified by inter-site comparisons and time series or cluster analyses (i.e., comparing concentrations among days). In some case studies, the concentration spikes were tracked to nearby source influences using meteorological information. For example, high concentrations of 1,3-butadiene collected hourly at a site in the Houston Ship Channel were associated with winds from the direction of industrial emissions.

- Metadata (e.g., satellite photos, site photos, emission inventory information) are needed to identify spikes as possibly real or not representative of broader urban or regional concentrations because of local source influences.

- In the next phase of analyses, guidelines will be developed to further address this question.

What have we learned about the frequency of air toxics monitoring in order to capture the true story on exposure?

- Sampling frequency is a function of the length of time separating sample collection (e.g., daily versus a sample every third day) and the duration of the sample (e.g., 1-hr versus

24-hr averages). For exposure assessment, one metric used is the annual average concentration. For several air toxics, archive data were used to show that site-specific AA concentrations can be estimated with 10% to 15% relative error using every third to every sixth day sampling (for data sets with concentrations mostly above the MDL). More frequent sampling is needed for higher concentration (source-oriented) sites to obtain similar relative error.

- For an investigation of sample duration, the archive data were limited to only a few air toxics including selected carbonyl compounds and VOCs. The data showed that, for benzene at urban sites and for formaldehyde, 1-hr to 3-hr resolution should be considered if diurnal variation of these compounds is a factor important to the understanding of their impact on health.

What are appropriate methods for producing daily, annual, or city-wide averages?

- While case studies performed using the archive and pilot city data provided a basis for formulating guidelines on the preparation of daily, annual, and city-wide averages, the guidelines were not specifically developed. The next phase of analyses will address this question in more detail. It is possible that different methods will be appropriate for different species. For example, carbon tetrachloride is very long-lived in the atmosphere, has had few legal emission sources in the past several years, and exhibits relatively low spatial and temporal variability—less stringent data completeness rules may apply to this air toxic than for a pollutant with more spatial and temporal variability.

How do we quantify uncertainty in the data analysis results?

- One measure of uncertainty was investigated by identifying the confidence in AA computations for the 18 HAPs by plotting percent CV versus sample size and concentration.

- The variability in the ambient concentrations due to spatial, temporal, sampling, and analytical variations was also investigated and documented.

- Depending on the analysis objective, quantification of uncertainty may not be necessary (or possible) for species with a large percentage of measurements below MDL. In those cases, only a qualitative answer may be possible.

What do air toxics data say about the effectiveness of various control programs?

How effective have mobile source controls been in reducing exposure to toxics?

- As discussed earlier, benzene reductions in reformulated gasoline have been directly linked to ambient benzene reductions.

How effective have maximum achievable control technology (MACT) standards been in reducing exposure to toxics?

- One of the case studies performed with the archive data involved changes in ambient concentrations before and after controls were implemented on a lead smelter in Missouri in the mid 1990s. Approximately a 90% reduction in the annual geometric mean lead

TSP concentration was observed after the controls were put in place. Another case study showed that ambient lead concentrations were significantly lower after the closure of a battery recycling plant.

- Both examples illustrated the usefulness of the archive data to investigate control effects; however, information (metadata) was needed from local officials that is not available in the data base.

Should other non-toxic species measurements be added to toxics monitoring sites to aid source apportionment?

- The addition of other species, either measured as part of the toxics program, or measured as part of other programs collocated with the toxics monitors, is useful to both validation and source apportionment efforts as previously discussed.

Have emissions controls reduced background concentrations?

- The 2004 analyses will address this question for carbon tetrachloride and perchloroethylene, at a minimum.

What effect have various community assessment projects had in reducing exposure to air toxics?

- Case studies performed using the historical database showed reductions in lead TSP concentrations as a result of plant closures (Tennessee) and the installation of control measures (Missouri). Thus, if community assessment project information (metadata) is available near sites with air toxics data, the database can be used to investigate potential air quality changes.

How has implementation of ozone and PM controls reduced air toxics levels (and vice versa)?

- This question was not addressed but will be considered in the 2004 analyses.

In Phases I and II of the air toxics data analyses, many policy-relevant questions were addressed in the broader context of the design of a national network but only limited portions of the available data set were used. Additional analyses are needed to refine the answers and to address the remaining questions. Phase III analyses will focus on these questions using an expanded view of the historical and pilot city databases and build on the Phase I and II analyses discussed here.

REFERENCES

Battelle Memorial Institute and Sonoma Technology, Inc. (2003) Phase II air toxics monitoring data: analyses and network design recommendations. Final technical report prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL by Battelle Memorial Institute, Columbus, OH, and Sonoma Technology, Inc., Petaluma, CA, December.

Bortnick S., Coutant B., Holdren M., Stetzer S., Holdcraft J., House L., Pivetz T., and Main H. (2001) Air toxics monitoring data: Analyses and network design recommendations. Report prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA and Battelle Memorial Institute, Columbus, OH, October.

Bortnick S.M., Coutant B.W., and Biddle B.M. (2003) Estimate background concentrations for the national-scale air toxics assessment. Final technical report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Battelle, Columbus, OH, Contract No. 68-D-02-061, Work Assignment 1-03, June.

Hafner H.R., Wheeler N.J., and Roberts P.T. (2004) Analysis of air toxics monitoring data. Work plan prepared for Lake Michigan Air Directors Consortium, Des Plaines, IL, by Sonoma Technology, Inc., Petaluma, CA, STI-903555-2442-WP2, January.

Main H.H., Roberts P.T., and Reiss R. (1998) Analysis of photochemical assessment monitoring station (PAMS) data to evaluate a reformulated gasoline (RFG) effect. Report prepared for the U.S. Environmental Protection Agency, Office of Mobile Sources, Fuels and Energy Division, Washington, DC by Sonoma Technology, Inc., Santa Rosa, CA, STI-997350-1774-FR2, April. Available on the Internet at <http://www.epa.gov/oar/oaqps/pams/rfg_oms.pdf>.

Main H.H. and Roberts P.T. (2000) PAMS Data Analysis Workbook: Illustrating the use of PAMS data to support ozone control programs. Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900243-1987-FWB, September.

Main H.H. and Roberts P.T. (2001) $PM_{2.5}$ Data Analysis Workbook. Draft workbook prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900242-1988-DWB, February.

Main H.H. (2002) Evaluation of control program effectiveness. Technical memorandum prepared for the Northeast States for Coordinated Air Use Management, Boston, MA, by Sonoma Technology, Inc., Petaluma, CA, STI-902071-2253-TM, September.