INTERLABORATORY QUALITY ASSURANCE FOR PHASE II OF THE NORTHERN CONTAMINANTS PROGRAM (1998-2003)

Yvonne D. Stokker

National Laboratory for Environmental Testing
National Water Research Institute
Burlington, Ontario L7R 4A6

January 2003

NWRI Contribution No. 03-~~xxx~~ 001

ABSTRACT

The quality, reliability and comparability of measurement data for various contaminants in Arctic samples were assessed over the period from 1998 to 2003 by means of a series of intercomparison studies. These studies addressed the analysis of heavy metals, methylmercury, organotins and persistent organic pollutants (POPs) such as the organochlorinated pesticides (OCs), PCBs and toxaphene using standard solutions, sediments, certified reference materials (CRMs) and a variety of natural biotic tissues from Northern Canada as test materials. The heavy metal analyses steadily improved over the course of the program and remain strongly metal and concentration dependent. Considerable confidence can be placed in the data generated for arsenic, cadmium, copper, mercury, selenium and zinc, while aluminum, chromium and nickel data were less reliable. Methylmercury and total organic mercury measurements were consistently reliable and comparable among the NCP participating laboratories, while some non-NCP participants demonstrated difficulties with these analyses. OC and PCB measurements were generally more accurate and comparable at analyte concentrations greater than 1 ng.g-1. It appeared that the laboratories' OC and PCB calibration standards were reliable but that some losses and/or contamination problems were occurring during the sample preparation steps, particularly in the high lipid tissues. Toxaphene data should continue to be scrutinized carefully, particularly for total toxaphene measurements. Toxaphene congener analyses were relatively more accurate and comparable among the NCP laboratories. Dibutyltin and tributyltin measurements were reliable for both sediment and biotic test samples, but monobutyltin data were highly dependent on the quality of the calibration standard. Future NCP intercomparisons should be conducted to ensure that acceptable levels of precision and accuracy are generated for the measurement of OCs, PCBs, toxaphene, heavy metals, and methylmercury. As more facilities begin to address the analysis of polybrominated diphenylethers (PBDEs), polychlorinated naphthalenes (PCNs) and fluorinated compounds, additional intercomparisons addressing these emerging issues should be conducted.

RÉSUMÉ

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Interlaboratory Quality Assurance for Phase II of the Northern Contaminants Program (1998-2003)

INTRODUCTION

The Northern Contaminants Program (NCP), like all research and monitoring programs, requires an ongoing quality assurance (QA) program. Such a program provides assurance to managers of the quality, reliability, and comparability of measurement results being generated for their research projects. At the same time, it should also meet the diverse quality assurance/quality control (QA/QC) needs of the researchers and analysts by providing them with appropriate diagnostic tools for their analyses and offering guidance and support toward corrective measures, if needed.

The main objective of the NCP QA program is to provide information to its science managers on the overall quality of the NCP's measurement data to assist them in making informed decisions on the sources of contaminants and their effects on the Arctic environment and on human health. As a result, these decision-makers would be assured that their contributions toward establishing international agreements and controls to protect the health of the Arctic ecosystem and its inhabitants are based on a scientifically sound database of information.

A second objective of the QA program is to assure research managers of the NCP-funded studies of the quality, reliability, and comparability of measurement results produced by

laboratories contributing data to their research. This is primarily achieved by conducting intercomparison exercises on various contaminants of concern, such as persistent organic pollutants (POPs) and heavy metals.

This report provides information on the overall quality of the NCP's analytical data from 1998 to 2003, and focuses on issues such as accuracy, precision, bias, and data comparability among laboratories. The results of a series of intercomparison studies are summarized and recommendations are made for future QA activities. Data quality issues associated with the human health studies are addressed in "Quality Assurance Aspects of the Human Health Studies" in the Human Health chapter of the Canadian Arctic Contaminants Assessment Report II (Van Oostdam et al., 2003).

BACKGROUND

PHASE I of the NCP (1993–1998) devoted considerable attention to identifying contaminants and pathways, monitoring trends, and establishing controls on substances of concern in the Arctic environment. The focus of PHASE II was shifted more toward immediate human health and safety issues associated with these contaminants in traditionally harvested foods for Northern people. At the onset of Phase II of the NCP, it became clear that other external QA programs were already competently addressing the quality of data for many NCP parameters and some matrices of interest. In many cases, this was achieved through accreditation to international standards for environmental, nuclear, and health/hygiene laboratories. Two surveys were therefore conducted in 1998

to assess and prioritize the data quality needs of the NCP, particularly in terms of analytes and matrices of interest.

The first survey identified the organizations that were contributing scientific data to the NCP, reviewed their analytical programs and capabilities, and assessed their existing quality control measures (Stokker and Gomes, 1999a). Table 1 lists the measurement laboratories that participated in the NCP Interlaboratory QA Program over the last 5 years. Approximately half of these laboratories contributed measurement data to various NCP research studies during Phase II of the program. Many of these facilities were either accredited by, or in compliance with, regulations established by external agencies, indicating that a sound quality management system was in place. These agencies included the Standards Council of Canada (SCC), the Canadian Association for Environmental Analytical Laboratories (CAEAL), the American Industrial Hygiene Association (AIHA), and the Atomic Energy Control Board (AECB). It was determined from the survey that existing quality control measures, including participation in external interlaboratory programs, adequately assured the data quality for determinations of trace metals, organochlorine pesticides (OCs), polychlorinated biphenyls (PCBs), and nutrients in water, for the measurement of radionuclides, and for analyses of human health tissues (e.g., blood, urine, hair). Due to the lack of available intercomparison programs employing appropriate matrices with suitable target analyte concentrations, however, the following were determined to be the highest priority for annual NCP intercomparisons:

(a) heavy metals in sediment and biota;

(b) methylmercury in biotic tissues;

(c) organochlorine pesticides (OCs) and polychlorinated biphenyls (PCBs) in biotic and abiotic samples; and

(d) toxaphene in biotic tissues.

Table 1. Laboratories participating in the Northern Contaminants QA Program

Organization	Interlaboratory study participation
Analytical Service Laboratories Ltd. Vancouver, British Columbia	*trace metals*
Atomic Energy of Canada Limited Whiteshell Labs Pinawa, Manitoba	*trace metals in sediments*
Aurora Laboratory Services Ltd. Vancouver, British Columbia	*trace metals*
Axys Analytical Services Ltd. Sidney, British Columbia	*OC/PCBs, toxaphene*
Bureau of Chemical Safety, Food Research Division Health Canada, Ottawa, Ontario	*OC/PCBs*
Centre de Toxicologie du Québec Ste-Foy, Québec	*trace metals, OC/PCBs, toxaphene*
Centre for Indigenous Peoples' Nutrition and Environment Ste-Anne de Bellevue, Québec	*trace metals and methylmercury, OC/PCBs, toxaphene*
Enviro-Test Laboratories Edmonton, Alberta	*trace metals and methylmercury, OC/PCBs*
First Nations and Inuit Health Branch Health Canada, Ottawa, Ontario	*trace metals, OC/PCBs*
Flett Research Ltd. Winnipeg, Manitoba	*trace metals and methylmercury*
Freshwater Institute Fisheries and Oceans Canada, Winnipeg, Manitoba	*trace metals and methylmercury, OC/PCBs, toxaphene*
Frontier Geosciences Inc. Seattle, Washington, USA	*trace metals and methylmercury*
Great Lakes Institute for Environmental Research University of Windsor, Windsor, Ontario	*trace metals*
Great Lakes Laboratory for Aquatic Sciences Fisheries and Oceans Canada, Burlington, Ontario	*OC/PCBs, toxaphene*
Great Lakes Science Center United States Geological Survey, Ann Arbor, Michigan, USA	*toxaphene*
Indiana University Bloomington, Indiana, USA	*toxaphene*
Institut des sciences de la mer de Rimouski (ISMER) Université du Québec à Rimouski, Rimouski, Québec	*organotins*
Institut Maurice-Lamontagne Fisheries and Oceans Canada, Mont-Joli, Québec	*OC/PCBs, toxaphene*
Institute of Ocean Sciences Fisheries and Oceans Canada, Sidney, British Columbia	*OC/PCBs*
McMaster University Hamilton, Ontario	*trace metals*

Organization	Interlaboratory study participation
National Laboratory for Environmental Testing Environment Canada, Burlington, Ontario	*trace metals and methylmercury, OC/PCBs, toxaphene*
National Water Research Institute Environment Canada, Burlington, Ontario	*trace metals, toxaphene, organotins*
National Wildlife Research Centre Environment Canada, Hull, Québec	*trace metals, OC/PCBs*
Northwest Atlantic Fisheries Centre Fisheries and Oceans Canada, St. John's, Newfoundland	*organotins*
Norwest Soil Research Ltd. Surrey, British Columbia	*trace metals*
Nunavik Research Centre Makivik Corporation, Kuujjuaq, Québec	*trace metals*
Ontario Ministry of Environment Toronto, Ontario	*OC/PCBs, toxaphene*
Philip Analytical Services Bedford, Nova Scotia	*trace metals*
School of Public Health, Environment and Occupational Health University of Minnesota, Minneapolis, Minnesota, USA	*toxaphene*
SRC Analytical Saskatoon, Saskatchewan	*trace metals*
Taiga Environmental Laboratory Yellowknife, Northwest Territories	*trace metals*
Université du Québec (INRS-Eau) Sainte-Foy, Québec	*trace metals*
University of Guelph Guelph, Ontario	*trace metals*
Wastewater Technology Centre Burlington, Ontario	*OC/PCBs*
Wellington Laboratories Guelph, Ontario	*OC/PCBs*

The second, concurrent survey reviewed the suitability of a number of external interlaboratory QA studies that would complement or could even substitute for NCP-run intercomparison studies. These external intercomparisons were of particular interest for the less commonly analyzed parameters such as organotins, polynuclear aromatic hydrocarbons (PAHs), dioxins and furans, as well as chemicals of emerging concern such as the brominated flame retardants (BFRs) and polychlorinated naphthalenes (PCNs).

Table 2. External interlaboratory programs in which NCP laboratories participate

Program & Organization	Analytes	Matrices
CAEAL Proficiency Testing Program Canadian Assoc. for Environmental Analytical Laboratories, Ottawa, ON	Trace metals, Nutrients, OCs/PCBs, PAHs	Water, sediment/soil, oil, air filters
CFIA Mercury Quality Assurance Program Canadian Food Inspection Agency, Winnipeg, MB	Hg	Fish
CFIA Fish Check Sample Program Canadian Food Inspection Agency, Mississauga, ON	PCBs	Fish
CTQ Interlaboratory Comparison Programs Centre de Toxicologie du Quebec, Ste-Foy, QC	Trace metals	Blood, serum, urine
InterLaB WatR, InterLaB Soil, Environmental Resource Associates (**ERA**), USA	Trace metals, Nutrients, OCs/PCBS, PAHs, Dioxins/Furans	Water, soil (fortified materials only)
Integrated Atmospheric Deposition Network (**IADN**) Environment Canada, Downsview, ON	Trace metals, OCs/PCBs, PAHs	Standard solutions, water
IAEA-AQCS (Analytical QC Services) Intercomparison Exercises International Atomic Energy Agency, Austria	Radionuclides, Trace metals, OCs/PCBs, Methylmercury,	Sediment, fish, lichen, algae, coral sand, soil, water
Hair Mercury Quality Control Program First Nations and Inuit Health Branch (**FNIHB**), Ottawa, ON	Hg	Human hair
NOAA Intercomparison Exercise for Trace Metals in Marine Sediments and Biological Tissues, National Oceanographic and Atmospheric Admin., USA	Trace metals	Sediment
NOAA Intercomparison Exercise Program for Organic Contaminants in the Marine Environment, Maryland, USA	OCs/PCBs, PAHs	Sediment, mussels, fish
NWRI Ecosystem Quality Assurance Program National Water Research Institute, Burlington, ON	Trace metals, Major Ions & Nutrients, Hg	Water
QUASIMEME International Laboratory Performance Studies and Development Exercises, Scotland	Trace metals, PAHs, Nutrients, Chlorinated organics, VOCs, PCBs, Dioxins/furans, Toxaphene, Organotins	Seawater, soil, sediment, fish, shellfish, standard solutions and extracts

Table 2 lists several of these external interlaboratory programs whose regularly scheduled studies were recommended as supporting the data quality needs of the NCP QA Program (Stokker and Gomes, 1999b). Recommendations were also made for participation in certain international development exercises that were unique, one-time-only intercomparisons.

Each year, prior to scheduling the new intercomparison studies, a list of the analytes and matrices being measured in the research studies was compiled along with updated information on the NCP laboratories and their performance in previous studies. This allowed the most appropriate series of intercomparisons for the QA program to be designed in order to accommodate shifting priorities and new matrices, as well as to focus on data quality issues revealed in previous studies.

INTERLABORATORY COMPARISON STUDIES

Table 3 provides a list of interlaboratory studies that were conducted under Phase II of the NCP. The studies were designed to identify sources of measurement uncertainty and variation of analytical results. By serving as a diagnostic tool for the participating analysts, these studies also provided a means of continually improving the measurement process. As outlined above, the studies were specifically designed to address emerging issues and a variety of matrices, and to monitor QA issues that were revealed in earlier intercomparisons. For the OC/PCB and toxaphene studies, the test samples became progressively more complex in each intercomparison.

Table 3. Interlaboratory studies conducted from 1998 to 2002 during Phase II of the Northern Contaminants Program.

Study No.	Target Analytes	Test Samples
NCP II-1	heavy metals	Great Lakes sediment CRMs
NCP II-2	heavy metals methylmercury total organic mercury	whole lake trout (Northern Québec) ringed seal muscle (Northern Québec) mussel homogenate CRM, fish muscle CRM
NCP II-3	21 organochlorines 30 PCB congeners 4 coplanar PCBs	standard solutions of OCs, PCBs, and coplanar PCBs whole lake trout (Great Lakes) dried mussels CRM
NCP II-4	total toxaphene toxaphene congeners	technical toxaphene solutions (Hercules standard) mix of 13 toxaphene congeners lipid-free burbot liver (Yukon, Northwest Territories)
NCP II-5	heavy metals methylmercury total organic mercury	Narwhal muktuk (Nunavut) land-locked char fillets (Nunavut) burbot liver (Yukon, Northwest Territories)
NCP II-6	organotins (specifically MBT, DBT, and TBT)	standard solutions dried sediments, dried fish CRM, dried mussel CRM
NCP II-7	heavy metals methylmercury total organic mercury	caribou liver (Northwest Territories) ringed seal liver and kidney (Baffin Bay) Greenland shark muscle
NCP II-8	22 organochlorines 30 PCB congeners 12 WHO PCBs	mixed OC/PCB standard solution solution of 12 WHO PCB congeners polar bear blubber (Alaska), ringed seal blubber (Baffin Bay) Lake Ontario Coho salmon, Lake Superior siscowet
NCP II-9	total toxaphene homologue totals toxaphene congeners	technical toxaphene, mixture of 15 toxaphene congeners Lake Superior siscowet beluga whale blubber (Nunavut)
NCP II-10 (proposed for 2002)	heavy metals methylmercury	polar bear liver, seabird tissue, whole walleye sediment
NCP II-11 (proposed for 2002)	toxaphene	technical toxaphene and congener mix solutions seal blubber, burbot liver extract

DATA ASSESSMENT TECHNIQUES

Interlaboratory data were assessed for accuracy, precision, and bias, and where possible, Z-scores were calculated (Thompson and Wood, 1993). To evaluate accuracy, the submitted results for standard solutions and matrix test samples were compared to their target concentrations. In some studies, certified reference materials (CRMs) were used with known target concentrations, while in other cases, the real matrix samples had target values determined by consensus from data submitted in the study. Where sufficient data were received, accuracy and comparability were also evaluated by Z-scores after rejecting outliers using a Grubb's test at the 5% significance level (Grubbs, 1969). Precision was evaluated by means of replicate analyses, by percent difference on blind duplicate samples, or by graphical interpretation of Youden Pair plots (Youden, 1959, 1960).

Bias, which is an indication of systematic error, was determined by a modified Youden-ranking procedure (Youden, 1962, 1963, 1969), or by graphical interpretation of Youden Pair plots. A set of results was said to be biased when the set exhibited a tendency to be consistently higher or lower than the results from the other participants. This ranking procedure and the criteria employed in testing for bias have about one chance in 20 of deeming a set of results biased, when in fact it was not (i.e., α=0.05). When bias was found, it suggested the presence of a systematic error that should be identified and corrected by the laboratory.

Z-scores were reported only for analytes for which sufficient data were received in the study. They were calculated using the original QUASIMEME-specified target standard deviation of 12.5% for real matrix test materials (Cofino and Wells, 1994). This target variability represents what would be achievable if the participating analysts were able to distinguish between two samples that differed by 50% in concentration. For the trace metal studies, a graphical presentation of "% satisfactory Z-scores" showed which laboratories generated the most reliable data with the least number of outlying results relative to their peers (Wells et al., 1997).

A complete data summary was provided at the close of each study as a diagnostic tool for the participants to apply corrective action, as needed. Where possible, the data review also included a comparison of the overall study results and conclusions of the results of other similar external intercomparisons. Therefore, each study also provided a snapshot of data quality to the science managers of the NCP along with an overview of the capabilities and comparability of the NCP laboratories conducting these measurements.

RESULTS OF THE INTERCOMPARISON STUDIES

In all intercomparisons, participants were asked to use their own in-house methodologies. In order to include as many laboratories as possible, both standard solutions and real matrix samples were used as check samples. By comparing the results submitted on standard solutions, the performance of abiotic and biotic measurement laboratories could be compared because their differences due to the analysis of unfamiliar matrices were

minimized. Most of the real matrix samples used as test materials in these studies were biotic in nature, reflecting the emphasis on traditional foods of the North in Phase II of the NCP. Although several laboratories that do not generate NCP measurement results were welcomed into the intercomparison program, the remainder of this discussion will focus primarily on the performance of the NCP laboratories.

Toxic heavy metals in sediment

Several NCP Phase I research studies measured loadings of metals such as mercury, cadmium, and lead in Arctic and sub-Arctic lakes (Lockhart et al., 1995, Barrie et al., 1997). Additional studies during Phase II continued to address temporal trends and spatial distribution of mercury and other metals in sediment cores of eastern and western Arctic lakes (Lockhart, 2000, Cheam et al., 2001). Therefore, the first heavy metal intercomparison, NCP II-1, was conducted to assess the ability of the NCP-funded laboratories to measure heavy metals in sediments. Four freeze-dried reference sediments, including one paired set of blind duplicates, were used as the test materials. The participating laboratories submitted results for at least 10 heavy metals, and in some cases, up to 23 metals.

A comparison to the reference and certified reference values of the test sediments showed that accurate data were being generated for most metals in sediment. However, several laboratories employed methods that did not include the use of hydrofluoric acid (HF), which is known to yield lower recoveries for many of the silicate minerals (Cook et al., 1997). This was particularly evident in this study for aluminum, chromium, thallium, and

vanadium. For most of the remaining metals, the data were quite comparable between laboratories, showing interlaboratory coefficients of variation of less than 25%. Accuracy and comparability in this study are shown graphically in Figure 1, which plots the mean % recovery for each metal in each of the four sediment CRMs. The error bars mark the highest and lowest % recovery of each metal in the sediment test samples. It can be seen from Figure 1 that the most accurate and comparable data were generated for cadmium, cobalt, copper, iron, manganese, nickel, lead and zinc.

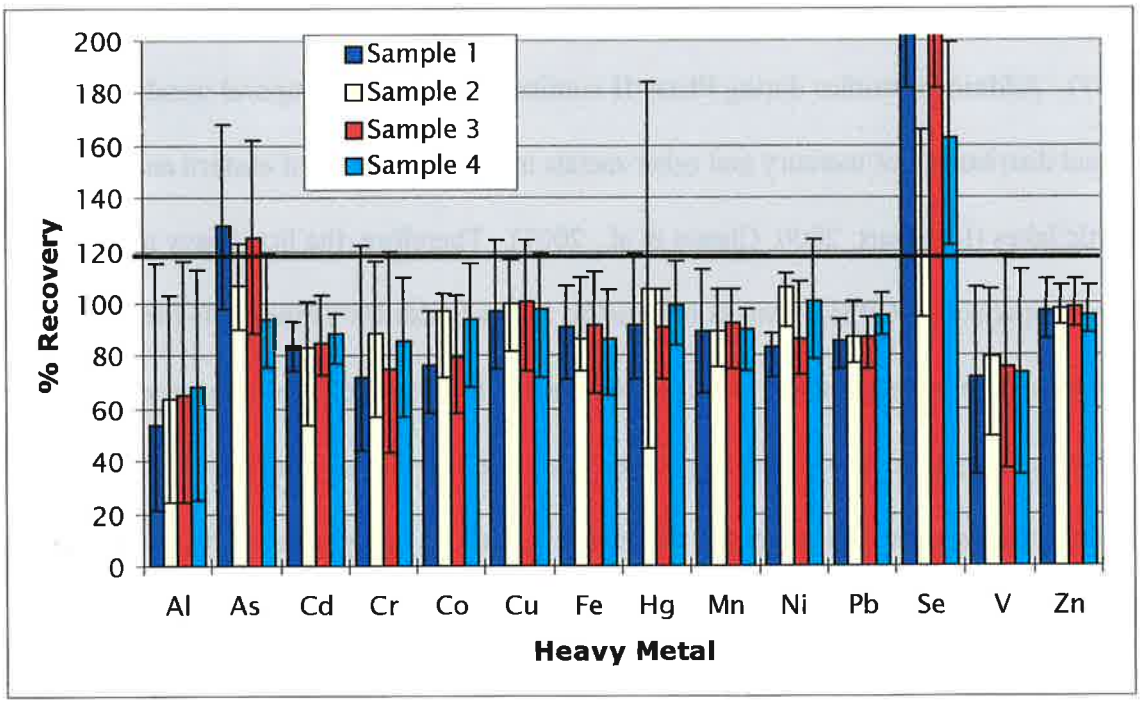

Figure 1. Mean % recovery of heavy metals from sediments in Study NCP II-1. Error bars mark the full range of recoveries by the study participants.

Intralaboratory precision, which was calculated as the percent difference between results for the blind duplicate samples, was very good for all but one of the participants. On the other hand, each participating laboratory demonstrated bias for at least one metal in this

study. This suggested that systematic errors were present and should be addressed by each participant in this program. Using the results for 11 metals and a target standard deviation of s=12.5%, all participants achieved better than 70% satisfactory Z-scores (i.e., $Z \leq 2$) on their analyses of heavy metals in sediments.

Heavy metals and methylmercury in biotic samples

Several research studies conducted during Phase I of the NCP revealed alarming trends of mercury levels in the environment as well as in many traditional food species, such as fish, birds, and marine mammals (Barrie et al., 1997). In addition, contamination of marine biota by heavy metals showed significant regional differences (Wagemann et al., 1996) and additional monitoring was recommended to address specific information gaps. Therefore, annual intercomparisons (studies NCP II-2, NCP II-5, and NCP II-7) were conducted to assess the measurement capabilities of NCP-funded laboratories to analyze biotic tissues for heavy metals such as mercury, cadmium, and lead, as well as for methylmercury.

In order to assess accuracy, the first biotic intercomparison for heavy metals included several internationally certified reference materials (CRMs). Many of the submitted results were well within the 95% tolerance limits listed for the CRMs' reference values and nearly all data were within ±25% of the certified reference values. The mean % recoveries ranged from 92 to 101% for all metals except aluminum (84%), chromium (72%), and nickel (71%). At methylmercury concentrations greater than 0.4 µg/g, most laboratories achieved recoveries that were also within ±25% of the reference values.

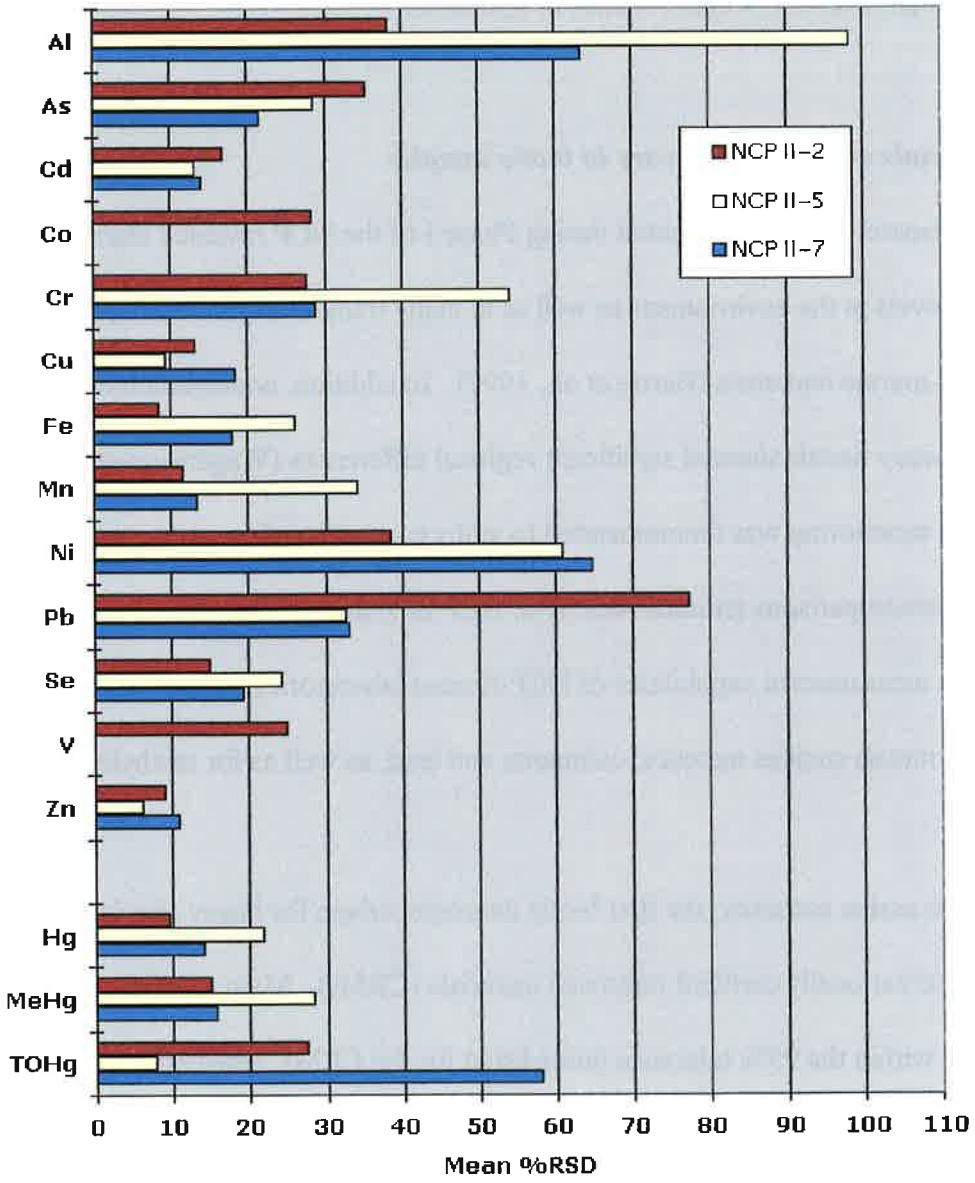

Figure 2. Relative standard deviations (%RSDs) of heavy metals and methylmercury for three intercomparisons on Arctic biota samples. The test samples for NCP II-5 (muktuk, char, and burbot liver) had very low concentrations of most metals.

As listed in Table 3, the remaining test samples in studies NCP II-2, -5, and -7 were homogenized biota samples from Northern Canada. The naturally low levels of the heavy metals in the Arctic biota caused a wider spread to the data than that found for the sediment samples of NCP II-1.

Nevertheless, for several metals, the interlaboratory coefficient of variation was better than 30% in all three studies. This is illustrated in Figure 2, which compares the mean relative standard deviation for each metal in the three studies. The higher variation in NCP II-5 is likely due to the additional challenges of homogenizing and digesting the muktuk, dealing with the high lipid content of the muktuk and burbot liver, and the very low concentration levels in the char. Figure 2 also shows that the least comparable results were generated for aluminum, chromium, nickel, and lead. Intralaboratory precision for the heavy metals, as determined by triplicate analyses of the seal muscle and whole lake trout samples in NCP II-2 and the caribou liver sample in NCP II-7, was generally very good for all NCP laboratories.

With few exceptions, the NCP laboratories generated excellent methylmercury results in terms of both accuracy (relative to the CRM reference values) and comparability among laboratories (low coefficient of variation), despite the use of very different extraction and analysis procedures at each facility. For total organic mercury, the data were also fairly agreeable between laboratories, as seen by interlaboratory coefficients of variation of less than 25%.

Figure 3 illustrates the precision and accuracy demonstrated by each laboratory by plotting the intralaboratory mean and standard deviation for methylmercury and total organic mercury for each participant on three test samples. (Laboratories B, E, F, and K do not contribute organic mercury data to the NCP program.) Figure 3 also shows that, despite the very low levels in the caribou liver, those laboratories that were able to measure methylmercury or total organic mercury were precise and comparable.

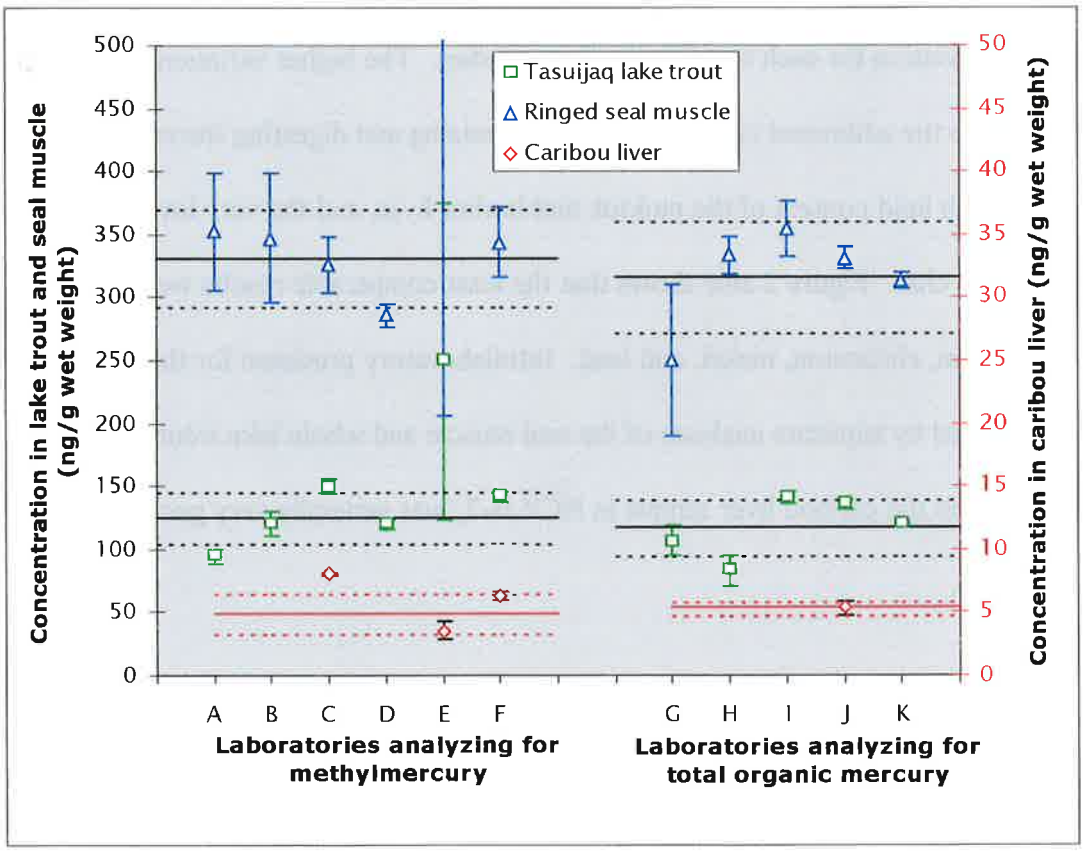

Figure 3. Precision and accuracy for methylmercury and total organic mercury. Intralaboratory mean values for methylmercury (six laboratories) and for total organic mercury (five laboratories) plotted on the Y-axes. The respective standard deviations of the means of triplicate analyses are shown as error bars. Horizontal lines represent the interlaboratory means ± the interlaboratory standard deviations. The caribou liver sample had levels too low for most laboratories to measure.

As a result of the very low levels of the target analytes in many of the samples, there were several outliers in each study, and consequently, several unsatisfactory Z-scores (i.e. $Z \geq 3$). However, most participants achieved better than 70% satisfactory Z-scores. In each study, only a few participants exhibited any metal-specific bias. This is a considerable improvement over the trace metals in sediment study, in which all participants exhibited systematic errors.

The results of these three heavy metal/methylmercury intercomparisons indicate that considerable confidence can be placed in the data being generated for most of the key toxic heavy metals. Aluminum, chromium, and nickel data, however, remain less reliable as the data were significantly less comparable over a wide concentration range. Methylmercury and total organic mercury results remain very reliable for the NCP measurement community. Monitoring of trace metals and methylmercury analyses should continue to be done on an annual basis, incorporating test materials with different concentration levels and varied digestion, analytical, and instrumental challenges.

OCs and PCBs

At the onset of this QA program, there were some concerns about the comparability of PCB measurements, particularly when individual researchers analyzed for and reported data for different sets of PCB congeners. To address this issue, PCB congener concentrations reported in the literature for marine or Arctic samples were compiled in order to tabulate the reported levels for various matrices of interest to the NCP studies. From these data, a list of 30 PCB congeners was recommended for future NCP work,

based on their toxicity, frequency of occurrence, and concentration levels relative to the (reported) total PCB levels (Stokker, 2000).

This list of congeners is provided in Table 4, along with comparisons to similar lists put forward by other national and international environmental monitoring programs including:

- Arctic Monitoring and Assessment Programme (AMAP) (Murray, 1999);
- International Council for the Exploration of the Seas (ICES) (ICES, 2001);
- International Atmospheric Deposition Network (IADN) (Cussion, 1993);
- National Oceanic and Atmospheric Administration (NOAA, 1993);
- Gulfwatch Monitoring Program (Gulf of Maine Council on the Marine Environment, 1998a and 1998b, Chase et al., 1998);
- Quality Assurance of Information for Marine Environmental Monitoring in Europe (QUASIMEME) (de Boer and Wells, 1997);
- The Quebec Ministry of the Environment (Lévesque and Moore, 1998); and
- The Canadian Shellfish Guidelines (Dumouchel and Hennigar, 1995).

In 2001, the subset of PCB congeners of toxicological importance was expanded from the four coplanar PCBs assessed in study NCP II-3 (i.e., PCB 77, 81, 126, and 169) to the 12 PCB congeners provided with a toxic equivalency factor (TEF) by the World Health Organization (WHO) (Van den Berg et al., 1998). Laboratory performance on the analysis of these 12 WHO PCB congeners was evaluated in NCP II-8.

Table 4. Comparison of NCP List of PCB congeners with those of other International Programs

PCB Congener No.	1999 NCP Minimum List	AMAP	ICES	IADN	NOAA	GULFWATCH mussels	QUASIMEME	QUEBEC MOE	Canadian Shellfish Guidelines
5		(x)		x		(x)			x
8	√	x/5		x	♦	♦/5			x
15				x		(x)			x
17								x	
18	√	x	x	x	♦	♦/15		x	x
28	√	♦/31	x	x	♦	♦	x	x	x
29				x		♦			x
31	√	(♦)		x			x	x	x
33								x	
44	√	x	x	x	♦	♦		x	x
49	√	x		x				x	
50						♦			
52	√	♦	x	x	♦	♦	x	x	x
66	√	(x)	x	x	♦/95	♦/95			x
70				x				x	
74	√			x				x	
77	B	x		x	B	♦			x
81	B			x					
82								x	
87	√	x		x		♦		x	x
90		(♦)		x	(x)	(x)			x
95		x/66		x	(x)	(x)		x	x
99	√	x		x				x	
101	√	♦/90	x	x	♦/90	♦/90	x	x	x
105	√	♦/132		x	♦	♦	x	x	x
110	√	x		x				x	x
118	√	x	x	x	♦	♦	x	x	x
126	B	x		x	B	♦			x
128	√	x		x	♦	♦		x	x
132		(♦)		x		(x)		x	x
138	√	♦/163	x	x	♦/163 /164	♦	x	x	x
146		x							
149	√	x		x				x	
151	√	x		x				x	
153	√	♦	x	x	♦	♦/132	x	x	x
154						♦			
156	√	x		x			x	x	
157		(x)							
158								x	
159					(x)				
163		(x)		x	(x)				x
164					(x)				x
169	B	x		x		♦		x	x
170	√	x	x		♦/190	♦/190		x	x
171								x	
177								x	
178									x
180	√	♦	x	x	♦	♦	x	x	x
182					(x)				
183	√	x		x				x	
187	√	x	x		♦/182 /159	♦		x	x
190					(x)	(x)			x
191								x	
194	√	x		x				x	
195	√	x		x	♦	♦/208		x	x
199								x	
201	√	x/157							
205								x	
206	√	x	x		♦	♦		x	x
208						(x)		x	x
209	√	x	x		♦	♦		x	x
Total # congeners:	30	7 essential 39 (total)	14	84	18-20	25	10	41	35

Program relevance:

√ - Recommended for NCP work
♦ - Essential
x - Recommended

B - indicates congeners of toxicological importance
(♦) - co-elutes with an essential congener
(x) - co-elutes with a recommended congener

During Phase II of the NCP, two intercomparison studies were conducted on the analysis of organochlorinated pesticides (OCs) and polychlorinated biphenyls (PCBs). Study NCP II-3 utilized separate check sample solutions for OCs and for PCBs, a Great Lakes fish and two biotic tissue CRMs, while study NCP II-8 employed a mixed OC/PCB standard solution and a variety of northern biota as test samples. (Refer to Table 3 for details). Injection-ready standard solutions were included as check samples in order to assess the quality of calibration solutions being used and to evaluate the separation and quantitation techniques for both OCs and PCBs without the confounding influence of matrix effects.

The results for the analysis of the 21 OCs in an iso-octane solution were extremely good. Most participants were within 25% of the design values, while the interlaboratory coefficient of variation for most parameters was better than 20%. For the analysis of PCB congeners in injection-ready solutions, most of the reported results were within 15% of their design values and interlaboratory comparability ranged from 7 to 40%. For the mixed OC/PCB solution used in study NCP II-8, the participants continued to produce very accurate and comparable results except for p,p'-DDD, o,p'-DDT, and p,p'-DDT. These three parameters had a number of outlying results reported by several participants, resulting in considerably higher coefficients of variation. With few exceptions, PCB congener results reported for the mixed standard solution were consistently within 20% of their design values and showed coefficients of variation up to 20% for the 30 target PCB congeners and up to 30% for the 12 WHO congeners. The widest variation was seen for congeners 66, 95, and 209. A close examination of these data suggests that the close elution of PCB66 and PCB95 gave rise to some difficulties in the correct identification of

these congeners for a few participants. Collectively, these results provide considerable confidence in the quality of OC and PCB calibration solutions being used by the NCP laboratories. In addition, this indicates that the participating laboratories were in control of the separation, identification, and quantification of individual OCs and PCB congeners.

The results for the tissue CRMs were considerably less accurate than for the standard solutions. This suggests that some analyte losses may be occurring during the extraction or cleanup of the biotic samples. For the certified fish homogenate and the certified mussel tissue, % recoveries of the OCs ranged from 31 to 137%. Figures 4a and 4b illustrate the accuracy and comparability of OCs in standard solutions compared to those in the biotic tissue CRMs. It is clear from these graphs that the interlaboratory means of the OC analytes in the fish and mussel CRMs were considerably less than their respective target values, represented by the diagonal line.

There was also less comparability among the participants for OC measurements in the biotic tissues than for the standard solutions, particularly where concentration levels were less than 10 ng/g. At concentrations greater than 10 ng/g, the OC coefficients of variation (CVs) generally ranged up to 30%; between 1 and 10 ng/g, the coefficients of variation ranged up to 50%; and at concentrations less than 1 ng/g, interlaboratory coefficients of variation were as high as 116%. Despite the more generous OC concentrations in the polar bear blubber, seal blubber, and siscowet fish tissue, these high lipid samples had interlaboratory CVs between 15 and 72%.

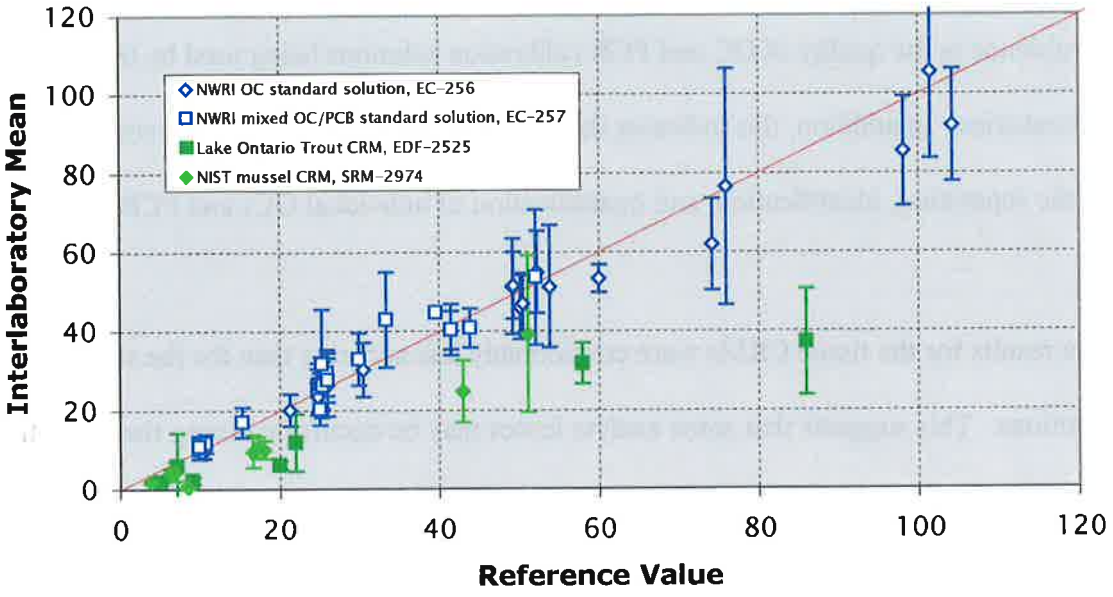

Figure 4a. Accuracy and comparability of OCs in standards (pg/µL) and biota (ng/g).

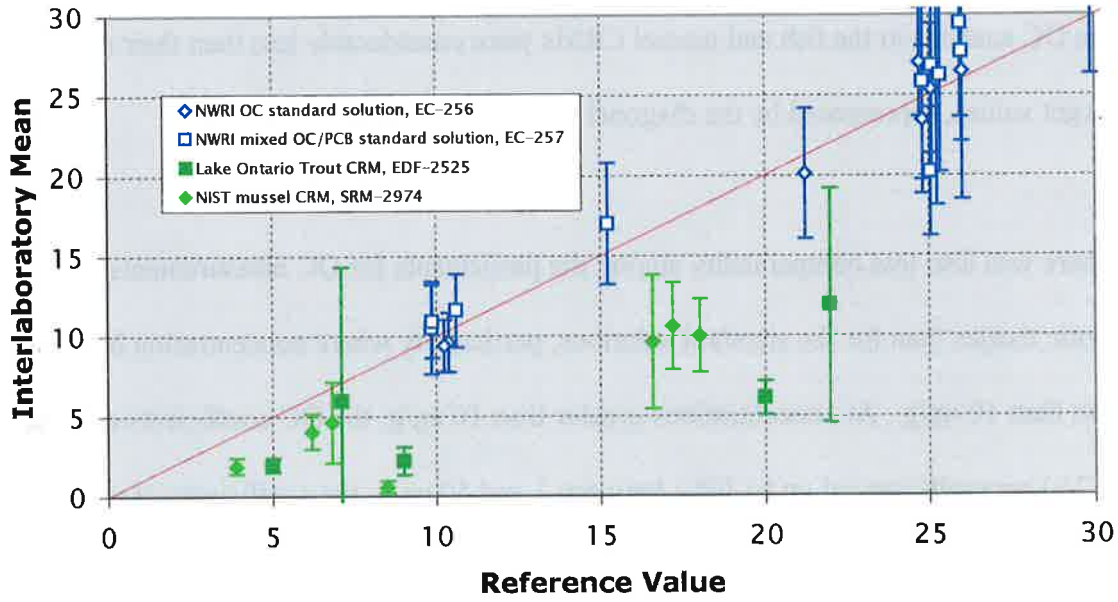

Figure 4b. An expanded view of Figure 4a at low concentration levels of OCs.

In figures 4a and 4b, error bars represent the standard deviations of the interlaboratory results. The diagonal lines indicate where the interlaboratory means would equal the reference values if 100% accuracy were achieved.

Accuracy of PCB congener analysis in the two certified tissues was much better than that for the OCs as the mean recoveries of the individual congeners ranged from 67 to 104%. For most congeners, the interlaboratory comparability was also good: coefficients of variation ranged from 10 to 45% in most of the tissue samples, even at levels approaching 1 ng/g. For the black guillemot liver, for which individual PCB concentrations were predominantly below 1 ng/g, some of the interlaboratory CVs ranged as high as 74%. Similar to the OC measurements, somewhat more variation was seen among the data for the high lipid polar bear and seal blubber samples.

In study NCP II-3, nine participants contributed results for the four coplanar PCBs in the injection-ready solution. The agreement to target (accuracy) and the agreement between laboratories (between-lab precision) were excellent for this test sample. For the three matrix samples in this study, fewer laboratories were able to quantify the considerably lower concentration levels of these parameters. Nevertheless, with the exception of PCB-81 results from one (non-NCP) laboratory, the coplanar PCB data for the biota samples were in close agreement with each other (better than 30% coefficient of variation for all samples). This is shown in Figure 5, where the interlaboratory coefficients of variation for the four coplanar PCBs are plotted for the two reference standard solutions and the three biotic samples in study NCP II-3.

Although additional results are still forthcoming for study NCP II-8 at the time of this writing, most of the data received so far for the 12 WHO PCB congeners are also very comparable among laboratories. Many of the interlaboratory coefficients of variation are

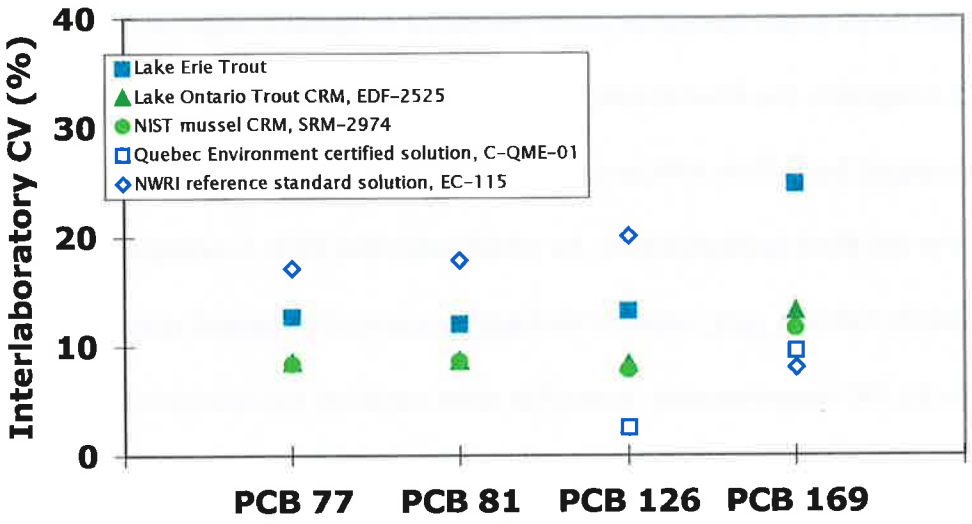

Figure 5. Comparability of coplanar PCBs in standard solutions and biota samples.

well below 40%. The most notable exceptions are for the black guillemot liver, which had PCB congener concentrations that were one to two orders of magnitude lower than in the other samples. The statistical results for the data already received are plotted in Figure 6 along with the nominal concentrations for each congener in the different tissues. It can also be seen from this graph that the data for PCB 77 in the two blubber samples had poor agreement among the laboratories.

In these studies, the OC, congener PCB, and toxic WHO PCB congener analyses were produced using a variety of extraction, cleanup, and instrumental procedures. Although the datasets were too small to draw any firm conclusions about the effect on data quality due to the differences between methodologies, it became apparent that there were

considerable differences in reported detection limits. The detection limits reported by the participants varied more than 100-fold for the OCs (from 0.01 to 5 ng/g) and ranged over four orders of magnitude for some of the PCB congeners (from 0.0001 to 1 ng/g). In many Arctic biota samples, methods with high detection limits could have a significant influence on data quality at low concentration levels.

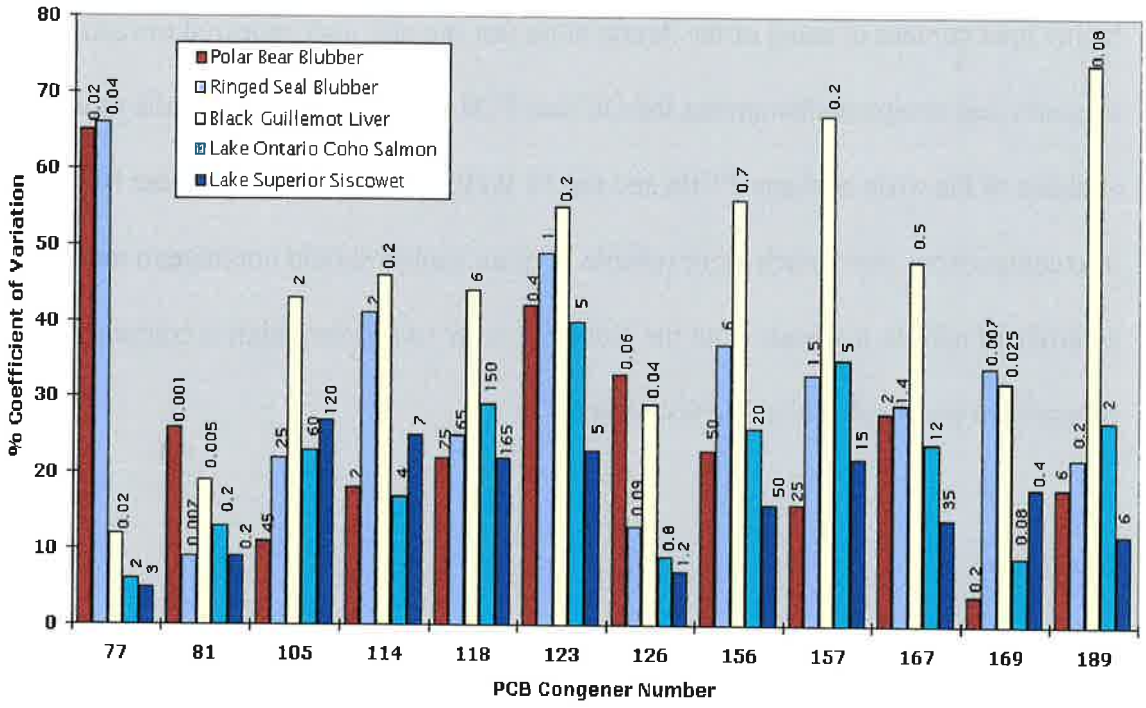

Figure 6. Comparability of WHO PCB congeners in biotic samples in Study NCP II-8. Nominal concentrations for each congener in the different tissues are noted above their respective bar in ng/g wet weight.

To summarize: the analyses of OCs and PCB congeners in injection-ready solutions were generally quite accurate and comparable. This provides considerable confidence in the quality of calibration solutions being used by the NCP laboratories. The analyses of these

same parameters in biotic tissue samples, however, were considerably less comparable, particularly for the OCs, where several extreme values contributed to between-lab variations exceeding 100%. This is in agreement with the findings of other international scientists conducting similar intercomparison studies: that currently available analytical methods for congener PCBs and OCs do not allow for the production of very accurate results when analyte concentrations are below 1 ng/g (de Boer and Wells, 1997). The higher lipid content of some of the Arctic biota test samples also appeared to reduce accuracy and comparability among the OC and PCB congener results. On the other hand, analyses of the toxic coplanar PCBs and the 12 WHO PCB congeners in these NCP intercomparisons were much more reliable. Future studies should continue to incorporate a variety of natural materials from the North, in order to accommodate a continuing progression of complexity in the test samples.

Toxaphene

The toxaphene methodology review conducted from 1999 to 2000 revealed a very diverse set of instrumental and quantitation techniques as well as many types and sources of calibration standards being used by the NCP measurement laboratories. Therefore, in order to eliminate differences in standards as the primary source of variation between laboratories, the first NCP toxaphene intercomparison included the provision of a calibration standard (purchased from Promochem GmbH) which the participants were to use as their calibration solution for the analysis of toxaphene congeners. The four check samples in this study included one test solution of 13 toxaphene congeners, two technical

toxaphene standards for total toxaphene analysis, and one lipid-free burbot liver extract for congener-specific and total toxaphene analysis.

The results for this first toxaphene intercomparison were very encouraging, particularly for the congener-specific analyses. The coefficients of variation for the individual (non co-eluting) toxaphene congeners in the blind test solution ranged from 7 to 25%, showing very good agreement between the ten laboratories. These results compare favourably with the European Research Project MATT where it was determined that "a CV value of about 20% is about the best of what can be obtained at the moment, even for more experienced laboratories" (de Boer et al., 1999). One concern in this study is the magnitude of some of the "false positives" (i.e., results reported for congeners not added to the test sample solution) reported by some laboratories. These are most likely due to system contamination at the laboratory or errors in peak identification by the analysts.

Not unexpectedly, congeners P26, P40, P41, P44, P50, and P62 dominated the toxaphene spectrum for the burbot liver extract. In addition, several participants reported comparable results for significant levels of congeners P38 and P42. Congeners P40 and P41 co-eluted for most analysts in this study and were therefore statistically assessed as one entity. There was good agreement between laboratories for all the above congeners as the interlaboratory coefficients of variation ranged from 18 to 49%. This is similar to the results seen in the QUASIMEME program where "CV values of 16 to 39% were obtained for the congeners P26, P50, and P62 in cleaned marine mammal and fish extracts" (QUASIMEME, 1998, de Boer et al., 1999).

There was less agreement between laboratories for total toxaphene, although this could partly be due to the provision of a calibration solution for the congener-specific analysis. The two technical toxaphene standards had interlaboratory coefficients of variation of 48% and 43%, and the data for total toxaphene in the burbot liver extract had a coefficient of variation of 56%. Figure 7 is a Youden Pairs plot (Youden, 1959 and 1960) of the total toxaphene data for the two technical toxaphene test samples. This graph clearly shows a bimodal distribution for both of the technical toxaphene standards. The data from one group of laboratories centred about the design values, while a larger group of laboratories reported total toxaphene results at about 50% of the design concentrations. As a result, the interlaboratory means and medians were significantly lower than the target values assigned by the supplier of this standard. Either most laboratories are underestimating their total toxaphene measurements or the reported concentration of this particular commercial standard was incorrectly identified. Figure 7 also demonstrates precision within the participating laboratories: the increasing perpendicular distance from the diagonal line is proportional to decreasing precision (Youden, 1959 and 1960). Therefore, despite the poor interlaboratory comparability, this plot of the two test samples as Youden Pairs revealed that the intralaboratory precision for total toxaphene was very good for all participants.

Our toxaphene methodology survey indicated that several researchers were quantifying total toxaphene against a Hercules technical toxaphene standard, while others employed technical toxaphene standards from the U.S. Environmental Protection Agency or other

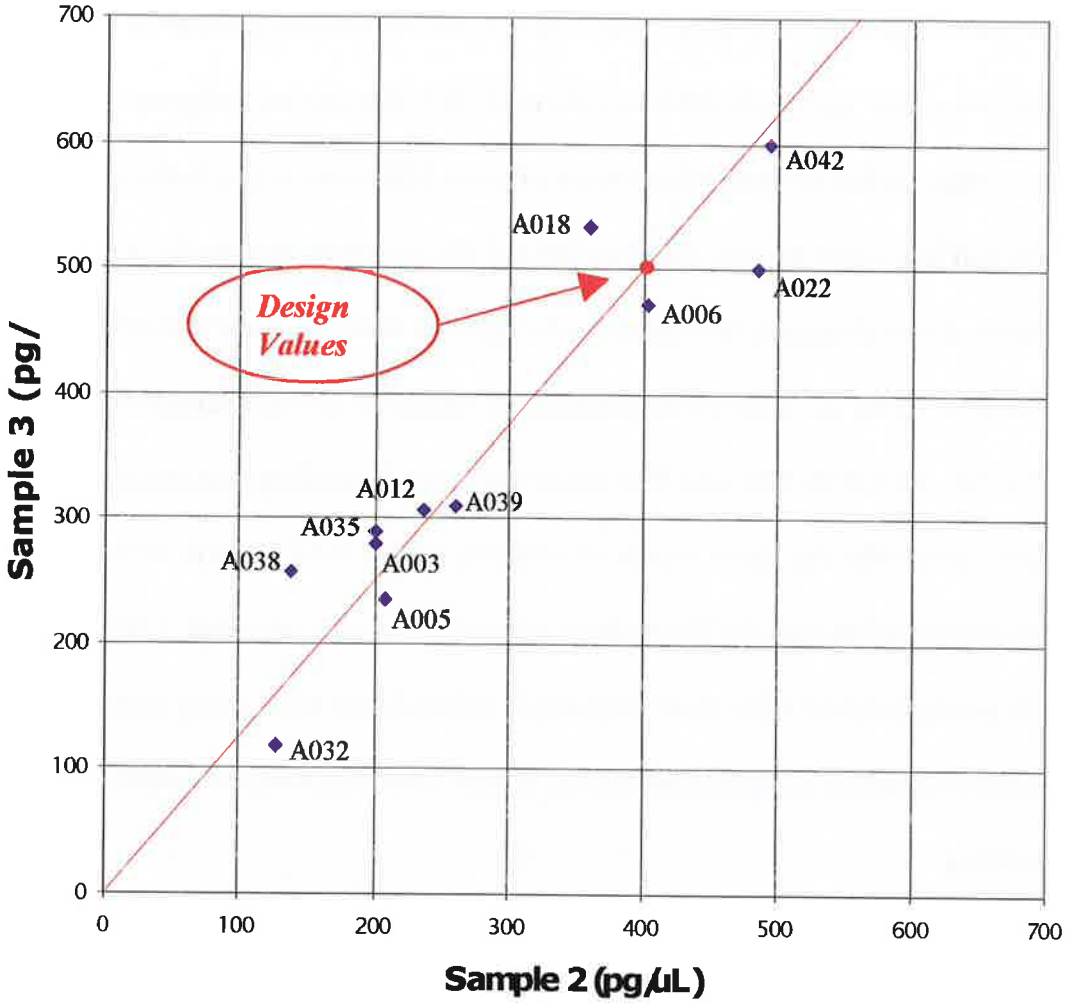

Figure 7. Youden Pairs plot of total toxaphene results for two Hercules standards. The perpendicular closeness of each data point to the diagonal line indicates acceptable intralaboratory precision for all participants.

(commercial) suppliers. Most calculated total toxaphene by using a single response factor against the sum of 4 to more than 40 peak areas, while two used multiple response factors for some or all of their total toxaphene calculations (Stokker, 2000).

As indicated by the wide-ranging recoveries against the Promochem calibration standard provided, there were significant differences between the participants' congener standards. Further work is needed to identify how much of these differences is due to the calibration standards and how much to peak identification and the quantitation procedures used by the analysts. Even among the laboratories who reported using the same standard (Dr. Ehrenstorfer's 'Parlar 22'), there was considerable variability between congeners P26, P38, P39, P41, P44, P56, P58, and P62, which could not be attributed to one particular laboratory. Generally, the Parlar source of standards tended to be equal to or somewhat higher in concentration than the Promochem calibration solution provided. Until this variation among in-house calibration standards is reduced, however, it may contribute significantly to a lack of comparability among sample results generated by different laboratories.

In the first intercomparison study, negative ion chemical ionization mass spectrometry (NICI-MS) was the predominant technique used for quantitation. Eight participants analyzed their samples by GC-MS (gas chromatography-mass spectrometry), two used Ion Trap GC-MS-MS, one combined the data from GC-MSD (mass selective detector) and GC-ECD (electron capture detection), and one used GC-ECD with a microbore column. For the individual congeners, most analysts reported detection limits of 0.1 to 5 pg/μL,

while one laboratory reported congener-specific detection limits of 0.01 to 0.1 pg/μL for their Ion Trap method. Detection limits for total toxaphene ranged from 5 to 100 pg/μL. Full details of the analytical and quantitation techniques can be found in the study report for NCP II-4.

The overall performance and comparability of the NCP laboratories in the first toxaphene intercomparison were very encouraging. However, there were also indications that significant differences remain between the laboratories, particularly for total toxaphene data. To ensure comparability with other international programs that address toxaphene, the selection of target parameters for future studies will continue to follow the recommendations put forth by AMAP (AMAP, 1998, pg 312) that

> "*Future monitoring should ... include determination of total toxaphene (by NIMS) for comparison with past work as well as measurements of specific chlorobornane congeners.*"

Therefore, the second toxaphene intercomparison (NCP II-9) is focusing on the separation and identification of individual congeners and the quantitation of total toxaphene. In order to include several American laboratories that do not routinely measure toxaphene congeners, homologue totals were also requested of the participants. The study samples include standard solutions to assess accuracy as well as more complex matrix samples: a beluga blubber sample and a high lipid fish homogenate prepared from a large Lake Superior siscowet. This study is still in progress at the time of this writing.

Radionuclides

Several NCP research projects continue to incorporate the analysis of radionuclides into their work each year. However, most of these measurements have been conducted at only two facilities, both of which routinely participate in more than one international program of radionuclide intercomparisons (Stokker, 2000). The first facility has a full QA program in compliance with the Atomic Energy Control Board requirements and ISO Guide 17025 (Standards Council of Canada, 2000). They also participate regularly in interlaboratory programs conducted by the U.S. Environmental Protection Agency (U.S. EPA) National Exposure Research Laboratory (NERL), Chalk River Labs of Atomic Energy of Canada Limited (AECL), and the World Health Organization's (WHO) International Reference Centre for Radioactivity in France. The second institution participates frequently in the intercomparisons offered by the U.S. EPA NERL and the International Atomic Energy Agency (IAEA) in Austria.

Stable Lead Isotopes

The quality and reliability of stable lead isotope measurements is a growing area of analytical concern. Past NCP studies have reported on these measurements (Lockhart et al., 2000) and they are becoming increasingly popular in global studies of climate change and contaminant source identification (Kurkjian et al., 2000, Murphy and Katz, 1998).

The measurement of environmental isotopes is a powerful tool that can be used to identify contamination sources and pathways, to investigate climatic changes, and to study environmental forensics (Murphy and Katz, 1998) and human exposure and metabolism

(Smith et al., 1996, Smith et al., 1998). This is because isotopes, both stable and radioactive, are preserved in various natural archives such as lake and ocean sediments, glaciers, polar ice caps, precipitation and oceans, and even trees. Ratios of stable isotopes are often uniquely associated with a geologic formation, while certain naturally occurring processes can concentrate the abundance of one type of isotope in one location as compared to another. Lead's isotopic abundance is one of the least reproducible because various isotopes are the final products of the radioactive decay of a number of heavy elements (Woolard et al., 1998). The subtle changes in the concentration or type of isotopes can be accurately measured and compared to reconstruct an accurate history or to provide evidence of anthropogenic inputs of contaminants, such as lead from leaded gasoline.

Until the mid- to late-1990s, most studies employing stable lead isotope tracer methods used thermal ionization mass spectrometry (TIMS) to measure isotope abundance. This technique provides excellent isotope measurement sensitivity and precision, but is expensive, labour-intensive, and not well suited for rapid throughput of large numbers of biological samples (Gwiazda et al., 1998). More recently, scientists have explored the use of ICP-QMS (inductively coupled plasma quadrupole mass spectrometry) to measure lead isotopic ratios in environmental work. Recognized limitations to ICP-MS measurements include the inability to measure very low levels of lead and particularly the low abundance isotope ^{204}Pb (Delves, 1999), a relatively noisy sample introduction, spectroscopic interferences, and not enough precision to adequately distinguish environmental sources based upon their isotopic composition (Sardella, 2000). In the case of lead, the isotope

ratios of contaminant environmental lead typically vary in a narrow range (e.g., by less than 3 to 7%). For ICP-QMS, the general precision of measurement is considered to be 2 to 3%, with a precision on isotope ratios of 0.1 to 1% (Sardella, 2000, Moens and Jakubowski, 1998). The use of high-resolution technology, such as the double-focusing mass spectrometer, which can greatly improve the sensitivity and precision, is now becoming more common. A single-collector double focusing magnetic sector ICP-MS can achieve precisions of 0.05 to 0.2%, and when a multi-collector detection system is used, precision can be further improved to a level comparable with that of the more costly TIMS (Moens and Jakubowski, 1998). Therefore, it is becoming increasingly important to assess the precision and accuracy of stable lead isotope data by considering the limitations of the instrumentation used (Smith, 2000).

At present, there are no known external quality assessment programs for stable lead isotope measurements. Few certified reference materials (CRMs) are known to be commercially available. The National Institute for Standards and Technology (NIST) has the following three lead (Pb) wire CRMs available:

SRM-981: Common Lead Isotopic Standard;

SRM-982: Equal-Atom Lead Isotopic Standard; and

SRM-983: Radiogenic Lead Isotopic Standard.

Each has certified isotopic compositions for ^{204}Pb, ^{206}Pb, ^{207}Pb, and ^{208}Pb and for the atomic abundance ratios of $^{204}Pb/^{206}Pb$, $^{207}Pb/^{206}Pb$, and $^{208}Pb/^{206}Pb$.

When stable lead isotope measurements have a greater role in the NCP research studies, it is recommended that method performance be carefully documented with all data. Furthermore, precision and accuracy of such analyses should be established and monitored internally by the laboratories with the above CRMs, and externally by the NCP or other external QA providers, in appropriate intercomparison studies.

Organotins

The toxic impact of tributyltin (TBT) on marine organisms has become well known (Chau et al., 1997) and its discovery in harbour sediments and snails from Norway, Iceland and Alaska, including some from remote regions, has made it a contaminant of increasing concern (AMAP, 1998). It is released into the environment by the leaching of TBT-based anti-fouling paints used on boats and ships. Because the data on TBT levels in the Arctic are limited, the following recommendations were made in the AMAP (Arctic Monitoring and Assessment Program) Assessment Report:

> *"Surveys of TBT in harbour sediments in the Arctic should be carried out to assess the extent of TBT contamination."* and
>
> *"The risk exists that TBT biomagnifies and analyses should also be made on bottom-feeding fish, waterfowl, and marine mammals, particularly as these are components of the diet of Arctic peoples."* (AMAP, 1998 p. 307).

Although few NCP research projects had proposed to incorporate the analysis of organotins in their work (Stokker, 2001), eight Canadian facilities were invited to participate in intercomparison study NCP II-6. The seven test samples included two

standard solutions, three dried sediments, and two dried biota samples. The analysts were asked to report on the concentration levels of monobutyltin trichloride (MBT), dibutyltin dichloride (DBT), and tributyltin chloride (TBT).

There was excellent agreement between laboratories on the sediment samples with slightly more variability on the test standard solutions. This was not unexpected, as these compounds are known to readily swap ligands when stored in mixed solutions. The test standard solutions contained DBT at identical concentration levels: very good precision was shown for the analysis of DBT in these blind duplicates. However, this study also brought into question the quality of some commercially available MBT standards that were being used by a few of the participating laboratories. As a follow-up to this discovery, direct comparative analyses on these standards were made and the MBT standards in question were discarded.

Chemicals of Emerging Concern

In addition to the more commonly analyzed chemicals of interest, a number of additional contaminants of emerging concern have been addressed in several recent NCP studies. Included among these analytes are polychlorinated naphthalenes (PCNs), short-chain chloroparaffins (SCCPs), haloacetic acids, polychlorinated diphenyl ethers (PCDEs), and brominated flame retardants (BFRs) (Bidleman et al., 2001). The measurement analyses for these contaminants are highly specialized and are carried out in only a few laboratories. Therefore, the intercomparisons conducted by the NCP QA Program have not addressed these contaminants. To ensure the data quality of these measurements, however,

continued interchange of samples and standards among the NCP laboratories is encouraged, as well as participation in external development exercises, such as those run by QUASIMEME.

Among the BFRs, temporal and spatial trends of polybrominated diphenyl ethers (PBDEs) have been the focus for several years of three Canadian laboratories (Stern and Ikonomou, 2001), one of which reports PBDE results to the NCP. From 1999 to 2000, these three laboratories successfully participated in a special international PBDE intercomparison study conducted by researchers in the Netherlands in collaboration with the Bromine Science and Environmental Forum (de Boer and Cofino, 2001). These same researchers have recently completed a second intercomparison as a development exercise under the auspices of QUASIMEME.

OVERALL ANALYTICAL DATA QUALITY WITHIN THE NCP QA PROGRAM

Each year throughout Phase II of the NCP, approximately 30 research projects with an analytical measurement component have received NCP funding. Most of the laboratories participating in the NCP QA Program also participate in a number of other national and international performance evaluation programs, including certification and accreditation programs. On the environmental side, the Canadian Association of Environmental Analytical Laboratories (CAEAL) has accredited more than half of the NCP laboratories for a number of water and/or sediment procedures. Similarly, the NCP laboratories

conducting radionuclide analyses were in compliance with national standards for these measurements.

The reliability of heavy metal analyses has steadily improved over the past four intercomparison studies and remains strongly metal-dependent. Most participants in the studies have demonstrated good comparability for the key heavy metals such as arsenic, cadmium, copper, mercury, and zinc. Aluminum, chromium, and nickel data are less comparable among laboratories. In general, accuracy and comparability have been good where metal concentrations are generous, but deteriorate at the very low levels measured in some Arctic tissues such as land-locked char, burbot liver, and ringed seal muscle. Although bias for various metals was evident among all the participants at the onset of this program, it has improved considerably in the last two trace metal studies, with very few participants now exhibiting any metal bias. Methylmercury and total organic mercury data have generally been reliable throughout Phase II, with the latter measurements being somewhat limited by their higher detection limits.

The first interlaboratory study on the analysis of OCs and PCB congeners in standard solutions and fish tissue (NCP II-3) showed that the laboratories were generally quite accurate and comparable in their analyses of OCs and PCBs in standard solutions but were considerably less comparable on the fish tissue samples, particularly for the OC measurements. Interlaboratory data for the coplanar PCBs 77, 81, 126, and 169 were very good, both in the standard solutions and in the matrix samples. Results from the second interlaboratory study (NCP II-8) showed similar results: the laboratories

generated more accurate and comparable results for PCB congeners than for OCs, and the results for standard solutions were significantly better than for the real matrix samples.

These results suggest that the laboratories have been using good quality calibration standards but may be experiencing some losses and/or contamination problems with their sample preparation steps, particularly for high lipid tissues. Despite these concerns, most of the OC and PCB data have been acceptable at more generous concentration levels and are less reliable at concentrations below 1 ng/g.

The results of the first toxaphene intercomparison study confirmed that the different calibration standards being used contributed significantly to the variability among laboratories. This was particularly evident for total toxaphene measurements, which showed a bimodal distribution in the results. Similar to the findings in other international toxaphene intercomparisons, there was good agreement among the participants for their measurements of the key biotic congeners. However, high false positives were one of the main concerns for the congener-specific analyses, and should continue to be monitored by the use of blind test mixtures of toxaphene congeners.

SUMMARY AND NEXT STEPS

NCP-funded research and monitoring studies involve the analysis of a wide variety of different contaminants at trace and ultra-trace levels in various matrices including air, snow, water, sediments, plants, fish, bird tissues and eggs, marine and terrestrial mammals,

and human tissue samples. It is such a diverse program that it becomes difficult to assess the matter of comparability of data among the different measurement laboratories and between individual projects.

The NCP intercomparison studies used a variety of (mostly biotic) test materials representative of the types of samples being analyzed in the NCP research studies. In many cases, standard solutions were also used as check samples in order to include several abiotic testing facilities. Although each interlaboratory data summary represents only a snapshot of the quality of measurements being generated at a particular time, several key findings have emerged from this QA program.

- Considerable confidence can be placed in the reliability of data generated for the toxic heavy metals arsenic, cadmium, copper, mercury, selenium, and zinc, while aluminum, chromium, and nickel data show more variability between laboratories.
- Methylmercury and total organic mercury data have been consistently reliable and comparable among the NCP measurement laboratories, while some participating laboratories, external to the NCP, have demonstrated problems with both accuracy and precision.
- The quality of the OC and PCB calibration solutions in the NCP measurement community are good, but some losses, particularly for the OCs, have been demonstrated on biotic samples
- Organochlorine pesticides (OCs) and polychlorinated biphenyls (PCBs) are generally more accurate and comparable at analyte concentrations greater than 1 $ng.g^{-1}$.

- Because of the diversity in standards and quantitation techniques, toxaphene data should be scrutinized carefully, particularly those of total toxaphene measurements.
- Toxaphene congener analysis in standards and in biotic samples has generally been accurate and comparable among the NCP laboratories.
- DBT and TBT measurements are reliable for both sediment and biotic samples, while MBT data are highly dependent on the quality of the calibration standard used.
- External QA programs, including the assessment of compliance to national and international standards, have supported the data quality of radionuclide measurements for the NCP research projects.

Ideally, the QA program would ensure the reliability and comparability of analytical results for all target contaminants in each matrix and species, as well as among the individual laboratories contributing the measurement data. Unfortunately, such a broad scope would be too costly. Consequently, the approach taken so far within the NCP QA Program has been to assess existing data quality measures in each measurement facility, make recommendations for participation in complementary external intercomparison programs, and then prioritize the remaining analytes and matrices. As a result, the current series of NCP intercomparison studies were designed and conducted to address these gaps in data quality assurance.

Future NCP studies should continue to ensure that acceptable levels of precision and accuracy are generated for the measurement of OCs, PCBs, toxaphene, heavy metals, and methylmercury. As more facilities begin to monitor and contribute data on new persistent contaminants such as the PBDEs and PCNs, additional intercomparisons addressing these emerging issues should be conducted. The ultimate goal is to provide assurance to NCP managers and scientists of a reliable and scientifically sound base for their research and monitoring programs of the North.

REFERENCES

Arctic Monitoring and Assessment Programme. 1998. *AMAP Assessment Report: Arctic Pollution Issues*. Oslo, Norway, 859 p.

Barrie, L., R. Macdonald, T. Bidleman, M. Diamond, D. Gregor, R. Semkin, W. Strachan, M. Alaee, S. Backus, M. Bewers, C. Gobeil, C. Halsall, J. Hoff, A. Li, L. Lockhart, D. Mackay, D. Muir, J. Pudykiewicz, K. Reimer, J. Smith, G. Stern, W. Schroeder, R. Wagemann, F. Wania, M. Yunker. 1997. Chapter 2. Sources, Occurrences and Pathways. In *Canadian Arctic Contaminants Assessment Report*. Indian and Northern Affairs Canada, Ottawa, ON, pp. 25–182.

Bidleman, T.F., M. Alaee, and G. Stern. 2001. New Persistent Chemicals in the Arctic Environment. In *Synopsis of Research Conducted under the 2000-2001 Northern Contaminants Program*. Indian and Northern Affairs Canada, Ottawa, ON, pp. 93–103.

Chase, M., P. Hennigar, J. Sowles, S. Jones, R. Crawford, G. Harding, J. Pederson, C. Krahforst, D. Taylor, and K. Coombs. 1998. Evaluation of Gulfwatch 1997. Seventh year of the Gulf of Maine Environmental Monitoring Plan. Gulf of Maine Council on the Marine Environment.

Chau, Y.K., R.J. Maguire, M. Brown, F. Yang, and S. Batchelor. 1997. Occurrence of organotin compounds in the Canadian aquatic environment five years after the regulation of antifouling uses of tributyltin. *Water Qual. Res. J. Canada,* 32: 453–521.

Cheam, V., D. Muir, and D. Halliwell. 2001. Spatial Trends in Loadings and Historical Inputs of Metals Inferred from Arctic Lake Sediment Cores. In *Synopsis of Research Conducted under the 2000-2001 Northern Contaminants Program.* Indian and Northern Affairs Canada, Ottawa, ON, pp. 110–117.

Cofino, W.P. and D.E. Wells. 1994. Design and Evaluation of the QUASIMEME Inter-Laboratory Performance Studies: A Test Case for Robust Statistics. *Mar. Poll. Bull.*, 29 (4–5): 149–158.

Cook, J.M., M.J. Gardner, A.H. Griffiths, M.A. Jessep, J.E. Ravenscroft and R. Yates. 1997. The Comparability of Sample Digestion Techniques for the Determination of Metals in Sediments. *Mar. Poll. Bull.,* 34 (8): pp. 637–644.

Cussion, Sylvia. 1993. IADN Parameter List for Interlaboratory Studies. Personal communication.

de Boer, J. and D.E. Wells. 1997. Chlorobiphenyls and Organochlorine Pesticides in Fish and Sediments --- Three Years of QUASIMEME Laboratory Performance Studies. *Mar. Poll. Bull.,* 35 (1–6): 52–63.

de Boer, J., J. Klungsoyr, G. Nesje, S. Meier, B. McHugh, E. Nixon, and G.G. Rimkus. 1999. MATT: Monitoring, analysis and toxicity of toxaphene — improvement of analytical methods. *Organohalogen Compounds,* 41: 569–573.

de Boer, J. and W. Cofino. 2001. First world-wide interlaboratory study on polybrominated diphenylethers (PBDEs). *Chemosphere*, 46: 625–634.

Delves, T. 1999. Valid analytical measurements in clinical applications of ICP-MS. *VAM Bulletin* 20: 16–21.

Dumouchel, F. and P. Hennigar. July 1995. Canadian Shellfish Contaminants Monitoring QA/QC Analytical Guidelines. Environment Canada.

Grubbs, Frank. 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 1(1): 1–21.

Gulf of Maine Council on the Marine Environment. 1998a. The First Five Years of Gulfwatch, 1991–1995: A review of the program and results.

Gulf of Maine Council on the Marine Environment. 1998b. Evaluation of Gulfwatch 1997: Sixth year of the Gulf of Maine Environmental Monitoring Plan.

Gwiazda, R., D. Woolard, and D. Smith. 1998. Improved lead isotope ratio measurements in environmental and biological samples with a double focusing magnetic sector inductively coupled mass spectrometer (ICP-MS). *J. Anal. At. Spectrom.*, 13: 1233–1238.

International Council for the Exploration of the Sea (ICES). ICES Environmental Data Centre. 2001. < http://www.ices.dk/datacentre/data_intro.asp >.

Kurkjian, R., C. Dunlap, and A. Flegal. 2000. Post-Industrial sources of lead in urban air: a case study using lead isotopes in Yerevan, Armenia. *Atmospheric Environment*. In press.

Lévesque, S. and S. Moore. 1998. A New Specific Congener Method for the Characterization of Environmental Samples Containing PCBs. In *Proceedings of the Enviro-Analysis '98 Conference.* Ministère de l'environnement et de la faune du Québec.

Lockhart, L, G. Stern, and P. Outridge. 2000. Mercury and other contaminants in a laminated sediment core from Devon Island. In *Synopsis of Research Conducted under the 1999-2000 Northern Contaminants Program.* Indian and Northern Affairs Canada, Ottawa, ON, pp. 104–112.

Lockhart, L., P. Wilkinson, B.N. Billeck, R.V. Hunt, R. Wagemann, and G.J. Brunskill. 1995. Current and historical inputs of mercury to high-latitude lakes in Canada and to Hudson Bay. *Water, Air and Soil Pollution,* 80(1–4): 603–610.

Moens, L. and N. Jakubowski. 1998. Double-focusing mass spectrometers in ICPMS. *Analytical Chemistry News and Features* 70(7): 251A–256A.

Murphy, B.L. and M.F. Katz. 1998. Environmental Forensics Development: Isotope Methods, *Environmental Claims Journal* 10(4): 135–150.

Murray, Janine. July 1999. List of POPs considered essential or recommended for AMAP. Personal communication.

National Oceanic and Atmospheric Administration (NOAA). 1993. Sampling and Analytical Methods of the National Status and Trends Program National Benthic Surveillance and Mussel Watch Projects 1984–1992. Volume 1.

QUASIMEME Laboratory Performance Studies Round 14, Exercise 379 DE-2. 1998. Aberdeen, UK.

Sardella, G. 2000. Head, Inorganic Analysis Laboratory, National Laboratory for Environmental Testing, Burlington, ON, personal communication.

Smith, Donald. 2000. Environmental Toxicology, University of California, Santa Cruz, CA, USA, personal communication.

Smith, D.R., R. P. Illustre, and J.D. Osterloh. 1998. Methodological considerations for the accurate determination of lead in human plasma and serum. *Am. J. Ind. Health*: 33(5) 430–438.

Smith, D.R., J.D. Osterloh, and A.R Flegal. 1996. Use of endogenous, stable lead isotopes to determine release of lead from the skeleton. *Environ. Health Perspectives* 104(1): 60–66.

Standards Council of Canada (SCC). March 2000. Standards Council of Canada CAN-P-4D, General requirements for the competence of testing and calibration laboratories. (ISO/IEC 17025:1999).

Stern, G.A. and M. Ikonomou. 2001. Temporal Trends of Organohalogen Compounds in Canadian Arctic Beluga. In *Synopsis of Research Conducted under the 2000-2001 Northern Contaminants Program*. Indian and Northern Affairs Canada, Ottawa, ON, pp. 237–242.

Stokker, Yvonne, D. and Glynn Gomes. 1999a. Summary report of the analytical programs and capabilities of laboratories and organizations that contribute measurement data to the Northern Contaminants Program. NWRI Contribution No. 99-302.

Stokker, Yvonne D. and Glynn Gomes. 1999b. Summary report on the availability and suitability of external national and international quality assurance programs pertinent to the Northern Contaminants Program. 1999b. NWRI Contribution No. 99-301.

Stokker, Yvonne D. 2000. Northern Contaminants Program Interlaboratory Quality Assurance Program for 1999-2000. In *Synopsis of Research Conducted under the 1999–2000 Northern Contaminants Program*. Indian and Northern Affairs Canada, Ottawa, ON, pp. 329–339.

Stokker, Yvonne D. 2001. Northern Contaminants Interlaboratory Quality Assurance Program for 2000-2001. In *Synopsis of Research Conducted under the 2000–2001 Northern Contaminants Program*. Indian and Northern Affairs Canada, Ottawa, ON, pp. 364–371.

Thompson, M. and R. Wood. 1993. International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories. *Journal of AOAC International* 76(4): 926-940.

Van den Berg, M., L. Birnbaum, A.T.C. Bosveld, B. Brunström, P. Cook, M. Feeley, J.P. Giesy, A. Hanberg, R. Hasegawa, S.W. Kennedy, T. Kubiak, J.C. Larsen, F.X.R. van Leeuwen, A.K.D. Liem, C. Nolt, R.E. Peterson, L. Poellinger, S. Safe, D. Schrenk, D. Tillitt, M. Tysklind, M. Younes, F. Waern and T. Zacharewski. 1998. Toxic Equivalency

Factors (TEFs) for PCBs, PCDDs, PCDFs for Humans and Wildlife, *Environmental Health Perspectives* 106(12): 775–792.

Van Oostdam, J., S. Donaldson, M. Feeley, and N. Tremblay, D. Arnold, P. Ayotte, G. Bondy, L. Chan, E. Dewailly, C. Furgal, U. Gill, S. Kalhok, H. Kuhnlein, E. Loring, G. Muckle, E. Myles, O. Receveur, Y. Stokker, and B. Tracy. Canadian Arctic Contaminants Assessment Report II: Toxic Substances in the Arctic and Associated Effects — Human Health. 2003. Indian and Northern Affairs Canada, Ottawa, ON.

Wagemann, R., S. Innes, and P.R. Richard. 1996. Overview and regional and temporal differences of heavy metals in Arctic whales and ringed seals in the Canadian Arctic. *Sci. of the Tot. Environ.*, 186: 41–66.

Wells, D.E., A. Aminot, J. de Boer, W. Cofino, D. Kirkwood and B. Pedersen. 1997. A Review of the Achievements of the EU Project 'QUASIMEME' 1993–1996. *Mar. Poll. Bull.*, 35(1–6): 3–17.

Woolard, D., R. Franks, and D. R. Smith. 1998. Inductively coupled plasma magnetic sector mass spectrometry method for stable lead isotope tracer studies. *J. Anal. Atom. Spectrom.*, 13: 1015–1019.

Youden, W. J. 1959. Graphical Diagnosis of Interlaboratory Test Results. *Ind. Qual. Control*, 15(11): 1–5.

Youden, W.J. 1960. The sample, the procedure, and the laboratory. *Anal. Chem.*, 32(13): 23A–37A.

Youden , W.J. 1962. Measurement Agreement Comparisons Among Standardizing Laboratories. In *Precision Measurement and Calibration, Statistical Concepts and Procedures*, National Bureau of Standards Special Publication 300, Volume 1, pp.133–137.

Youden, W.J. 1963. The Collaborative Test. *J. Assoc. Off. Agr. Chem.*, 46: 55–62.

Youden, W.J. Ranking Laboratories by Round-Robin Tests. 1969. In *Precision Measurement and Calibration, Statistical Concepts and Procedures*, National Bureau of Standards Special Publication 300, Volume 1, pp.165–169.