A STUDY TO DETERMINE CURRENCY

OF DSL QUANTITY DATA

FOR USE IN CATEGORIZATION

OF DSL SUBSTANCES

Edward Doyle and Heather Patterson

August 2001

Introduction

The *Canadian Environmental Protection Act* (CEPA) provides a legislative basis for addressing environmental pollution problems in Canada. As an approach to the management of toxic substances, provisions of CEPA ensure that no new substance is introduced into the Canadian marketplace without being assessed by the Ministers of the Environment and Health to determine whether it is toxic to the environment or human health as defined in the Act. The Domestic Substances List (DSL) is the sole basis for determining whether a substance is "existing" or "new" to Canada and includes substances that, between January 1, 1984, and December 31, 1986, were manufactured in or imported into Canada in a quantity of not less than 100 kg in any one calendar year or were in Canadian commerce or used for commercial manufacturing purposes in Canada. These existing substances do not require notification under the New Substances Notification Regulations of CEPA.

In order to compile the DSL, Canadian chemical manufacturers and importers were required by Environment Canada to report the quantity of any substance manufactured, imported or used in Canadian commerce during the calendar year 1986 using range codes that spanned more than six orders of magnitude. The reporting ranges were defined as follows:

A	<100 kg
B	100–1000 kg
C	1000–10 000 kg
D	10 000–100 000 kg
E	100 000–1 000 000 kg
F	1 000 000–10 000 000 kg
G	>10 000 000 kg

In addition, notifiers were required to report the use codes associated with each substance, and these codes were segregated into the following categories: special use (including research and development and site-limited intermediate), functional use and industrial sector. The DSL was published in the Canada Gazette, Part II, on May 4, 1994.

Under the 1999 revisions to the CEPA legislation, Health Canada is required to categorize the DSL substances on the basis of (a) greatest potential for exposure of individuals in Canada and (b) persistence and/or bioaccumulation and inherent toxicity to humans; and to carry out screening-level risk assessments on those substances meeting the above criteria. In determining the greatest potential for exposure of individuals in Canada, it is expected that quantity in commerce data will be used along with use codes. However, since the DSL data were collected in 1986, there is a concern that changes in supply and demand patterns of the substances since that time may impact on the use of these data for categorization purposes. This study was undertaken to examine the extent and direction of these changes and to determine if the original data available on the DSL are representative of the current supply and demand patterns.

Methods

Data Collection

In order to determine whether current supply and demand patterns are comparable to the original quantity in commerce data as reported for the DSL, Chemical Product Index (CPI) profiles were obtained from Camford Information Services (Camford) for a selected list of substances, including organic and inorganic compounds as well as polymers (see Table 1). These profiles provide both current and past supply and demand data for chemical substances in Canada and are compiled by Camford on a regular basis for their clients using a combination of both published trade data and industry interviews.

The primary reference source used by Camford for trade data is Statistics Canada, which publishes monthly bulletins that report imports and exports by commodity. These data are used as the starting point for calculating the import and export volumes reported in the profiles. In those cases where the data cannot be used directly, Camford will make adjustments to increase the accuracy of the estimates. For example, chemical substances are often shipped in solutions of varying concentrations, but Statistics Canada does not attempt to adjust its volumes to account for these variations in concentration. In these cases, it is necessary to adjust the published data in order to achieve a proper material supply/demand balance. Also, Statistics Canada frequently groups similar materials together into "basket" categories in order to maintain confidentiality in markets with very few participants. In these cases, Camford will combine trade data based on the total in the "basket" category with information from other sources, such as the U.S. Bureau of Census, to arrive at accurate estimates.

Camford also reviews Natural Resources Canada publications on the production of leading minerals and their annual reviews of the markets for minerals and metals and Industry Canada publications on specific industry sectors, although these reports are also based on Statistics Canada data. Production volumes reported by various industry associations, such as the Canadian Chemical Producers Association, Canadian Pulp and Paper Association, Pulp and Paper Research Institute of Canada, Canadian Plastics Industry Association, Society for the

Plastics Industry, Canadian Fertilizer Institute, The Chlorine Institute and the National Petroleum Refiners Association, are also reviewed. Finally, Camford also reviews operating statistics and market information provided by public companies in their annual reports.

Relevant data from studies conducted by Camford for government agencies or private corporations may also be incorporated into the CPI profile database, provided that the results are not confidential. These studies typically cover a range of chemical processing industries, including petrochemicals, fertilizers, pulp and paper, mining and smelting, paint and coatings, soaps and detergents and plastics. If necessary, Camford will also acquire market studies from other research organizations (e.g., Kline & Company, CHEMinfo Services, Chem Systems, Freedonia Group, SRI International) to fill gaps in its own information base.

Table 1: List of substances for which CPI profiles were obtained from Camford Information Services

Chemical	CAS RN	Type	Chemical	CAS RN	Type
ABS terpolymer	9003-56-9	Polymer	Methanol	67-56-1	Organic
Acetaldehyde	75-07-0	Organic	Methyl ethyl ketone	78-93-3	Organic
Acetic acid	64-19-7	Organic	Methyl isobutyl ketone	108-10-1	Organic
Acetone	67-64-1	Organic	Methyl methacrylate	80-62-6	Organic
Acrylonitrile	107-13-1	Organic	Methyl tert-butyl ether	1634-04-4	Organic
Aluminum chloride	7446-70-0	Inorganic	Methylamine	74-89-5	Organic
Aluminum sulfate	10043-01-3	Inorganic	Methylene chloride	75-09-2	Organic
Ammonium chloride	12125-02-9	Inorganic	Morpholine	110-91-8	Organic
Ammonium nitrate	6484-52-2	Inorganic	Nitrogen	7727-37-9	Inorganic
Ammonium phosphate	10124-31-9	Inorganic	Oleic acid	112-80-1	Organic
Antimony trioxide	1309-64-4	Inorganic	Pentachlorophenol	87-86-5	Organic
Benzene	71-43-2	Organic	Perchloroethylene	127-18-4	Organic
Benzoic acid	65-85-0	Organic	Phenol	108-95-2	Organic
Butadiene	106-99-0	Organic	Phenolic resin	9003-35-4	Polymer
Butanol	71-36-3	Organic	Phosphorus	7723-14-0	Inorganic
Butyraldehyde	123-72-8	Organic	Polyacrylamide	9003-05-8	Polymer
Calcium carbide	75-20-7	Inorganic	Polyacrylate	9003-01-4	Polymer
Calcium carbonate	471-34-1	Inorganic	Polyacrylonitrile	25014-41-9	Polymer
Calcium chloride	10043-52-4	Inorganic	Polyethylene terephthalate	9003-68-3	Polymer
Calcium hypochlorite	7778-54-3	Inorganic	Polyisobutylene	9003-27-4	Polymer
Carbon black	1333-86-4	Inorganic	Polypropylene	9003-07-0	Polymer
Carbon dioxide	124-38-9	Inorganic	Polystyrene	9003-53-6	Polymer
Carbon disulfide	75-15-0	Inorganic	Polyvinylacetate	9003-20-7	Polymer
Cellulose acetate	9004-35-7	Organic	Polyvinyl alcohol	9002-89-5	Polymer
Chlorine	7782-50-5	Inorganic	Polyvinyl chloride	9002-86-2	Polymer
Chloroacetic acid	79-11-8	Organic	Polyvinylidene chloride	9002-85-1	Polymer
Chlorobenzene	108-90-7	Organic	Propylene glycol	57-55-6	Organic
Chloroform	67-66-3	Organic	Propylene oxide	75-56-9	Organic
Choline chloride	67-48-1	Organic	Rosin	8050-09-7	Organic
Citric acid	77-92-9	Organic	Sodium bichromate	10588-01-9	Inorganic
Cyclohexane	110-82-7	Organic	Sodium borohydride	16940-66-2	Inorganic
2,4-Dichlorophenoxy-acetic acid	94-75-7	Organic	Sodium chlorate	7775-09-9	Inorganic

Chemical	CAS RN	Type	Chemical	CAS RN	Type
Dimethyl terephthalate	120-61-6	Organic	Sodium chloride	7647-14-5	Inorganic
Ethanol	64-17-5	Organic	Sodium cyanide	143-33-9	Inorganic
Ethanolamine	141-43-5	Organic	Sodium hydrosulfite	7775-14-6	Inorganic
Ethoxylated alcohol	68002-96-0	Organic	Sodium hydroxide	1310-73-2	Inorganic
Ethyl chloride	75-00-3	Organic	Sorbitol	50-70-4	Organic
Ethyl acetate	141-78-6	Organic	Starch	9005-25-8	Organic
Ethylbenzene	100-41-4	Organic	Stearic acid	57-11-4	Organic
Ethylene	74-85-1	Organic	Styrene	100-42-5	Organic
Ethylene dichloride	107-06-2	Organic	Styrene-butadiene latex	9003-55-8	Polymer
Ethylene glycol	107-21-1	Organic	Sulfuric acid	7664-93-9	Inorganic
Ethylene oxide	75-21-8	Organic	Talc	14807-96-6	Inorganic
Ethylene-propylene rubber	9010-79-1	Polymer	Tall oil	8002-26-4	Organic
2-Ethylhexanol	104-76-7	Organic	Toluene	108-88-3	Organic
Formaldehyde	50-00-0	Organic	Toluene diisocyanate	26471-62-5	Organic
Glycerin	56-81-5	Organic	1,1,1-Trichloroethane	71-55-6	Organic
Hexane	110-54-3	Organic	Trichloroethylene	79-01-6	Organic
Hydrogen	1333-74-0	Inorganic	Urea	57-13-6	Organic
Hydrogen fluoride	7664-39-3	Inorganic	Urea resin	9011-05-6	Polymer
Hydrogen peroxide	7722-84-1	Inorganic	Vinyl acetate	108-05-4	Organic
Isopropanol	67-63-0	Organic	Vinyl chloride	75-01-4	Organic
Kaolin	1332-58-7	Inorganic	Xylenes	1330-20-7	Organic
Lead chromate	7758-97-6	Inorganic	Zinc	7440-66-6	Inorganic
Melamine	108-78-1	Organic	Zinc oxide	1314-13-2	Inorganic

In addition to reviewing published information, Camford also conducts extensive telephone and personal interviews to gather additional information for compiling the profiles. Every chemical manufacturer is interviewed at least once, and more than one person may be interviewed. For example, the plant manager is typically consulted for capacity and production data, while the sales manager is queried for sales volumes and estimates of the total size of each market segment. Distributors are also interviewed, including both major distributors (e.g., Van Waters & Rogers, Canada Colors and Chemicals, HCI Canada, Multichem, Ashland Canada) and smaller firms that may hold a significant position in specific markets (e.g., Frank E. Dempsey & Sons, Mallinckrodt Canada, JLM Chemicals Canada). These distributors are often in the best position to estimate the size of major market segments for the demand pattern.

Finally, end users in each market segment are also interviewed to ascertain how the chemicals are used and to determine per unit requirements, which are then used to derive market demand. Since end users listed in one profile may be producers in another profile, interviews and research on related profiles are conducted at the same time in order to maintain numerical consistency between profiles.

Data Analysis

Each CPI profile consists of two sections denoting the annual supply and demand patterns for a substance. Information on the supply pattern included the nameplate capacity of the industry and

the volumes of both domestic production and imports. Demand pattern information included the volume of the substance used for specific applications or products, total domestic demand and the volume of exports. The profiles generally cover the period from 1986 to the late 1990s.

Graphs showing the change in total supply over the reporting period were plotted for each substance. In addition, the percent change in total supply for each substance between 1986 and the latest year for which data are available was calculated using the equation:

$$\frac{\text{(Supply in latest year available – supply in 1986)}}{\text{Supply in 1986}} \times 100 = \%\ \text{change in supply}$$

These data were then used to calculate the mean, median, minimum and maximum changes that occurred for the total group of 110 substances, as well as the standard deviation and 95th percentile.

The temporal change in the use patterns of each substance was plotted as the percentage of total disappearance (equal to total supply) attributed to each use in each year for which data were available. This provided information on the percentage of total supply being acquired for each use and allowed for visual assessment of the magnitude of changes in use patterns that occurred throughout the reporting period.

Finally, the Camford quantity data for 1986 were compared directly with historic DSL data (1986 quantity in commerce) for each of the 110 substances to determine the degree of correlation between the two data sets. Historic DSL quantity data were generated by Environment Canada by assuming the upper bound of the reported range codes for each of the substances.

Results and Discussion

Visual examination of the change in supply patterns for each of the substances shows that there is a wide variation in the annual supply quantities reported, with some substances showing only small changes in supply over time and others undergoing large increases or decreases in supply. In most cases, however, the actual difference in supply between the base year (1986) and the most recent year for which data are available is less than an order of magnitude. In fact, only seven of the substances showed changes in supply quantities greater than an order of magnitude from the quantities reported in 1986. Included in these seven substances are methyl tert-butyl ether (MTBE), whose supply increased 3600% from 1988 to 1995, and hydrogen peroxide (H_2O_2), whose supply increased 991% from 1986 to 1999.

The large increases in supply of these substances can probably be accounted for by the rapidly increasing use of MTBE as an oxygenate in reformulated gasoline to reduce emissions of volatile organic compounds and air toxics relative to conventional gasoline, and by the increased use of

hydrogen peroxide as an alternative to chlorine in pulp bleaching operations to eliminate the formation of dioxin. Since these large increases in supply are not considered to be representative of typical growth patterns in the Canadian chemical industry, these two substances may be considered to be outliers for statistical purposes.

The other five of the seven substances that showed greater than an order-of-magnitude change were phosphorus, 1,1,1-trichloroethane, chloroform, ethyl chloride and butyraldehyde. These substances all showed decreases in supply quantities exceeding an order of magnitude from 1986 base values. The percent changes for these substances ranged from -90.7% to -99.9%.

Since the original DSL quantities were reported simply as ranges spanning an order of magnitude, then based on the fact that the changes in supply for 103 of the 110 substances (~94%) over the reporting period are within an order of magnitude of the quantity reported in the base year, it can be concluded that most of the original DSL quantity data can probably be used as estimates for current supply quantities, at least within an order-of-magnitude accuracy.

In another approach to examining the changes in supply quantities over time, quantity range codes as used in reporting for the DSL were assigned to the Camford supply quantities for 1986 and the most recent year for which data were available for each of the substances. For example, if supply quantities for a substance were reported to be 3903 kt in 1986 and 9875 kt in 1999, in both cases these quantities would fall into range code F (1 000 000–10 000 000 kg), which would not be considered to be any change in the supply quantity for this substance over the reporting period.

An examination of the assigned range codes shows that for 72 of the 110 substances (65%), there was a change in range codes over the reporting period; 35 of the substances (32%) showed a change in range code of at least one level (e.g., D→E, D→C, etc.), while 3 of the substances (3%) showed changes of two levels (e.g., D→F, etc.). However, since only a small change in the actual supply quantity can trigger a change in the reported range code (e.g., 999(B)→1001(C)), this approach to characterizing the changes in supply is not considered to be as useful as examining the actual magnitude of the change itself, as was done above.

Statistical analysis of the change in supply quantities between 1986 and the most recent year for which data were provided, including mean, median, minimum and maximum values, standard deviation and 95th percentile for the group of substances (with and without outliers), is summarized in Table 2 below. Thus, the mean percent change in total supply was found to be 77.7%, with 95% of the substances showing a change of less than 262.6%. The high variability associated with the data is not unexpected given the wide range of substances and use patterns. If MTBE and H_2O_2 are removed from the above calculations, based on the fact that they appear to be statistical outliers, then the mean percent change drops to 36.7%, with 95% of the remaining 108 substances changing less than 186.1%. This mean percent change is consistent

with the 30% growth in total Canadian industrial chemical output reported for the 10-year period 1986–1996 (*Chemical and Engineering News*, June 23, 1997).

Table 2: Percent change in the total supply of 110 substances

	Mean	Median	Minimum	Maximum	Standard deviation	95th percentile	Comments
TOTAL	77.73	22.44	-99.93	3600	363.44	262.58	110 substances
less H_2O_2 and MTBE	36.66	21.41	-99.93	576.19	95.43	186.13	
Inorganics	49.12	16.23	-99.12	991.36	182.54	149.89	32 substances
less H_2O_2	18.72	14.46	-99.12	186.44	62.30	118.51	
Organics	89.88	21.41	-99.93	3600.00	464.26	257.54	62 substances
less MTBE	32.34	21.23	-99.93	576.19	102.07	159.38	
Polymers	87.88	75.17	-73.46	313.58	109.52	275.22	16 substances
<100 kt	46.66	21.15	-99.93	991.36	154.93	266.72	77 substances
less H_2O_2	33.54	20.10	-99.93	576.19	107.69	220.28	
>100 kt	139.03	29.87	-53.59	3600.00	588.31	201.63	33 substances
less MTBE	42.89	29.13	-53.59	262.43	65.24	167.89	

Table 2 also shows the results of the statistical analysis (with and without outliers) for the substances when grouped by class (organics, inorganics and polymers) and when grouped by volume (<100 kt and >100 kt). Again, removal of the two outliers has a significant impact on the results.

In order to assess the accuracy of the original DSL quantity in commerce data reported to Environment Canada by the industry, the quantity in commerce data were compared with the Camford supply data for 1986. It is assumed that the Camford data more accurately reflected quantities in commerce because of the more comprehensive data collection techniques utilized. MTBE was not included in this comparison because Camford did not have data for MTBE for 1986.

The percent difference between the DSL data and the Camford data for 71 of the 109 substances (i.e., ~65%) is negative, indicating that DSL quantities in commerce for these 71 substances have been under-reported when compared with the Camford data, and the extent to which they have been under-reported is very large in many cases. For example, percent differences for ammonium phosphate, butyraldehyde, dichlorophenoxyacetic acid, ethoxylated alcohol, polyacrylonitrile, polyvinylidene chloride, propylene glycol and sodium chloride range from 19 981% to ~156 000 000%. No explanation for these large differences is readily apparent.

Since the Camford data are expected to be a more accurate estimate of quantities in commerce for 1986, there might be some justification for transforming the original DSL data prior to using them for categorization of the DSL substances if there is good correlation

between the two data sets. When compared using a distribution-free, non-parametric statistical method (Spearman's Rank Correlation), a correlation coefficient of only 0.61 was obtained. Using a simple linear regression analysis, a correlation coefficient (r) of 0.70 was found after eight substances were first eliminated from the analysis because of extremely large differences between the DSL and the Camford quantities in commerce. These substances were ammonium phosphate, butyraldehyde, dichlorophenoxyacetic acid, ethoxylated alcohol, polyacrylonitrile, polyvinylidene chloride, propylene glycol and sodium chloride. In this case, the proportion of variance in the DSL data associated with, or in common with, the variance in the Camford data was approximately 49% (i.e., $r^2 = 0.49$). In order to normalize the data and stabilize the variance, the natural logarithm (ln) transformation was used for the Camford data and the square root (sqrt) transformation was used for the DSL data. Fitting the model to the transformed data, the model adequacy improved, with the correlation coefficient (r) increasing to 0.73, indicating that the proportion of variance in the sqrt(DSL) data associated with, or in common with, the variance in the ln(Camford) data is approximately 53% (i.e., $r^2 = 0.53$). The regression model is:

$$\ln(\text{Camford}) = 7.313 + 0.019[\text{sqrt}(\text{DSL})]$$

Although this regression model could be used to transform all remaining DSL import quantities prior to using them for categorization purposes, this approach is not recommended, since the regression explains only 53% of the variability between the DSL and Camford data sets. The use of such a transformation could be justified only if the two data sets were shown to be highly correlated (i.e., $r^2 > 0.9$).

Camford profiles provided temporal changes in the use patterns for the 110 substances as the percentage of total disappearance (equal to total supply). While DSL use information was reported using predefined functional use and/or industrial sector codes, Camford use pattern information was more descriptive and/or specific, so it is difficult to make direct comparisons between the two sets of use data. For example, DSL records show that a substance had the following reported functional use and industrial sector codes: 04 (adhesives, binders, sealants and fillers), 35 (polymer, component of an article), 36 (polymer, component of a formulation), 55 (article manufacture) and 86 (plastics); Camford reports the following uses for the same substance: pipes and fittings, automotive parts, major appliances, communications equipment, other (small appliances, tools, toys, etc.) and export sales. However, in spite of the differences in reporting terminology, most of the uses identified by Camford for this substance would likely fall under one or more of the original use codes reported for the DSL, and this is expected to be the case for most of the substances.

A total of 592 uses has been reported by Camford for the 110 substances; in 68 of these cases (~11%), the percentage of total disappearance drops to zero at some point during the reporting period, indicating that a particular substance was no longer used for that application. Some examples include the use of ammonium chloride in dry cells, the use of n-butanol in the

manufacture of butyl phthalates and glycol ethers and the use of styrene-butadiene latex in the manufacture of high-solid foam. However, in most cases, the substances continue to show the same use patterns over the entire reporting period, although the quantities (as a percentage of total disappearance) tend to show considerable annual variation, reflecting the growth or decline of various markets for a substance. These results indicate that for most of the 110 substances, the use patterns have not changed significantly since 1986. Based on this information, it is reasonable to assume that most of the functional use and industrial sector codes reported in 1986 for the DSL substances would likely still be applicable to current markets.

Conclusions

Examination of the Camford data shows that only 7 of the 110 substances examined in this report showed absolute changes in supply greater than a single order of magnitude over the reporting period (1986 to the most recent year for which data were provided). Because the original DSL quantities were reported as ranges spanning more than six orders of magnitude, it is reasonable to conclude that most of the 1986 DSL quantities can be used as relative estimates of current supply quantities, at least within an order-of-magnitude accuracy.

Statistical analysis of the Camford data shows that the mean percent change for the selected substances between 1986 and the most recent year for which data were available (36.7%) is in good agreement with a reported growth of 30% in overall Canadian industrial chemical output during the 10-year period 1986–1996. This implies that the 1986 DSL quantities may on average be only ~30% less than current supply quantities, and certainly within the order-of-magnitude accuracy mentioned above.

When compared with Camford supply data for 1986, many of the DSL substances (~65%) appear to have been under-reported, if it is assumed that the Camford data are more accurate than the DSL data in estimating quantities in commerce. Thus, the DSL quantities in most cases do not represent a worst-case scenario in terms of quantities in commerce.

Transformation of DSL quantities in commerce prior to using the data for categorization purposes and based on a linear regression model relating Camford quantities and DSL quantities for a subset of 101 substances is not recommended, because of the high variability between the two data sets ($r^2 = 0.53$).

Based on a visual comparison of use pattern information provided by Camford on the 110 substances, it appears that for most of the substances, the use patterns have not changed significantly since 1986. Furthermore, it is probably reasonable to expect that this same pattern would also apply to many of the DSL substances, so that functional use codes and industrial sector codes reported in 1986 will still be relevant for categorization purposes in most cases.

Acknowledgments

The authors wish to thank Leonora Marro of the Biostatistics Group, Health Canada, for assistance in carrying out the statistical analysis of the data.