

THE EXPOSURE DATA LANDSCAPE FOR MANUFACTURED CHEMICALS

Peter P. Egeghy^{a*}, Richard Judson^b, Sumit Gangwal^b, Shad Mosher^b,

Doris Smith^c, James Vail^c, Elaine A. Cohen Hubal^b

^a National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States

^b National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States

^c Senior Environmental Employment Program, National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States

* Corresponding author at 109 TW Alexander Dr., Mail Drop E205-04, Research Triangle Park, NC 27711, United States, Tel +1 919 541 4103, Fax +1 919 541 0905, email:

egeghy.peter@epa.gov.

ABSTRACT

The U.S. Environmental Protection Agency is developing chemical screening and prioritization programs to evaluate environmental chemicals for potential risk to human health in a rapid and efficient manner. As part of these efforts, it is important to catalog available information on chemical toxicity and exposure from widely dispersed sources. The main objective of this analysis is to define important aspects of the exposure space and to catalog the available exposure information for chemicals being considered for analysis as part of the U.S. EPA ToxCast™ screening and prioritization program. Publicly available exposure data have been extracted into ACToR (Aggregated Computational Toxicology Resource), which combines information for hundreds of thousands of chemicals from >600 public sources. We use data from ACToR to assess the exposure data landscape for environmental chemicals. Of the roughly 100,000 chemicals that have at least limited toxicity information available, less than one-fifth also have exposure information – and for most of these the information is of limited utility (e.g., production volume). Readily accessible data on concentrations in exposure-related media are only available for a much smaller fraction. Among these, the largest number of chemicals is measured in water with over 1150 unique compounds, followed by 788 substances measured in soil, and 670 in air. These small numbers clearly reflect a focus of resources on those substances previously identified as possibly posing a hazard to human health. Exposure to a much broader number of chemicals will need to be measured in order to fully realize the envisioned goal of using exposure information to guide toxicity testing.

KEY WORDS: Exposure, database, ACToR, risk assessment, environmental chemicals

ACRONYMS

ACToR	Aggregated Computational Toxicology Resource
ATSDR	Agency for Toxic Substances and Disease Registry
BCF	Bioconcentration Factor
CASRN	Chemical Abstract Services Registry Number
CDC	Centers for Disease Control and Prevention
DSL	Domestic Substances List
EAFUS	Everything Added to Food in the United States
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EPA	U.S. Environmental Protection Agency
EPISuite	Estimation Program Interface
EU	European Union
FDA	U.S. Food and Drug Administration
HEDS	Human Exposure Data System
HPV	High Production Volume
HSDB	Hazardous Substances Data Bank
HUD	U.S. Department of Housing and Urban Development
IRIS	Integrated Risk Information System
IUR	Inventory Update Reporting
MPV	Moderate Production Volume
NHANES	National Health and Nutrition Examination Survey
NIOSH	U.S. National Institute for Occupational Safety and Health
NLM	National Library of Medicine

OSHA	U.S. Occupational Safety and Health Administration
PPD	Pesticide Data Program
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SPIN	Substances in Products in the Nordic Countries database
TOXNET	Toxicology Data Network
TRA	Targeted Risk Assessment
TSCA	Toxic Substances Control Act
USDA	United States Department of Agriculture

1. INTRODUCTION

The chemical industry has grown rapidly since the middle of the 20th century. Substantial investment of capital into research and development and product diversification has produced a continuous stream of innovations and a constantly expanding portfolio of products (Murmman, 2003). Manufactured chemicals have become integrated into nearly all industrial processes, building materials, and commercial goods such as furnishings, clothing, electronic equipment, cleaning products, and cosmetics (Weschler, 2009; Wilson and Schwarzman, 2009). While the contributions of the chemical industry to economic growth and to improvements in life expectancy, comfort, and living conditions in industrialized societies are widely recognized, there are also concerns about the health consequences of ubiquitous exposure to synthetic chemicals (Wilson and Schwarzman, 2009).

Approximately 8 million chemical substances are commercially available (Chuprina et al., 2010). Nearly 100,000 chemicals have been inventoried in U.S. commerce. These include about 82,000 substances regulated under the Toxic Substances Control Act (TSCA) as well as 8600 food additives, 3400 cosmetic ingredients, 1800 pharmaceuticals, and 1000 pesticide active ingredients regulated under other federal statutes. Approximately 30,000 of these substances are believed to be in wide commercial use, marketed in volumes above 1 ton per year (Muir and Howard, 2006). In 2006, roughly 3500 Moderate Production Volume (MPV) chemicals were produced in the U.S. at between 25,000 and 1 million pounds per year (lb y^{-1}), and about 2750 High Production Volume (HPV) organic chemicals were produced in volumes of greater than 1 million lb y^{-1} . Chemicals manufactured and used at the same site, for which exposures to consumers and the general public are expected to be less than for those that are not site-limited,

make up approximately 77 % of the 27 trillion lbs reported in 2006. The majority of the top chemicals by volume are produced during the refining of petroleum (EPA, 2008).

Information regarding the risks posed to human health and the environment from the large number of chemicals in commercial use is limited and often inadequate. Although relatively few chemicals are thought to pose a significant risk to human health, the hazard and exposure data necessary to adequately assess risk are unavailable for the vast majority of chemicals in commerce, even among the HPV chemicals (Applegate and Baer, 2006; EPA 1998). Of the nearly 10,000 HPV and MPV chemicals, pesticide ingredients, and drinking water contaminants considered for the U.S. Environmental Protection Agency (EPA) ToxCast™ screening and prioritization program, high-quality toxicology evaluation (e.g., National Toxicology Program, Agency for Toxic Substances and Disease Registry (ATSDR), EPA's Integrated Risk Information System (IRIS)) was unavailable for about three-quarters, and even limited toxicity information is lacking for one-third (Judson et al., 2009). Little effort has been put toward consolidating and evaluating the availability of exposure data.

Humans may be exposed to chemicals throughout the chemical's lifecycle, that is, during manufacture, distribution, product use, and disposal. The vast majority of chemicals in wide commercial use are not measured in environmental media, and their environmental fate and potential for human exposure are unknown. Even for those chemicals that have been determined to be hazardous, we have limited exposure data, rarely sufficient to estimate exposures for specific population subgroups. For the majority of the contaminants of emerging concern, even the most basic information regarding the occurrence in the environment is unavailable (Muir and Howard, 2006).

Recognizing the critical need for exposure information to inform chemical design, evaluation and health risk management, the ExpoCast™ program has been initiated by the US EPA in collaboration with stakeholders. The aim of the program is to ensure that the required exposure science and computational tools are developed and ready to: (1) address global needs for rapid characterization of exposure potential arising from the manufacture and use of thousands of chemicals, and (2) meet challenges posed by new toxicity testing approaches and integrated evaluation strategies. An early focus of this research program is to improve public access to exposure information. The objective of this paper is to identify resources useful for exposure assessment required to prioritize chemicals in commerce for screening, targeted testing and risk assessment. We investigate key aspects of exposure space and review exposure-related data resources. We analyze publically available databases housing exposure-related information to catalog available exposure data and identify strengths and limitations of exposure data available for chemical risk management.

2. METHODS

In this article we restrict our analyses to those chemicals collected in ACToR (Aggregated Computational Toxicology Resource) (Judson et al., 2008), the comprehensive data management system developed by EPA to manage the large-scale datasets of ToxCast™ (<http://actor.epa.gov>). ACToR is comprised of several independent data repositories linked to a common database of chemical structures and properties, and to tools for development of predictive toxicology. As such, it is a database holding essentially all publicly available information on chemical identity, structure, physical–chemical properties, *in vitro* assay results, and *in vivo* toxicology data (Judson et al., 2009). All chemicals in ACToR have been run through EPA’s EPISuite™ set of programs (EPISuite 4.0) to calculate predicted values of a series of physicochemical properties.

In particular, bioconcentration factor (BCF) was estimated with BCFBAF v 3.0 using the regression-based method, and biodegradation potential with BIOWIN v 4.1 using the “biowin1” linear model. Queries of the ACToR database (database version actor_2010q4a, accesses March 1, 2011) were performed to illustrate the types and amount of exposure data publicly available. The ACToR system is implemented in MySQL and can be freely downloaded so that others can submit advanced queries or develop custom data-mining applications.

3. EXPOSURE SPACE

“Chemical space” is a region defined by a particular choice of descriptors of molecular properties (e.g., molecular mass, octanol-water partition coefficient) and the limits placed on them. Identification of regions of chemical space that define bioactive molecules has been instrumental to drug discovery (Dobson, 2004; Chuprina et al., 2010). To identify relevant information for collation and evaluation, key aspects of an analogous “exposure space” particularly relevant to risk evaluation and risk management were considered. These included physicochemical properties, product characteristics, emissions characteristics, pathways, and receptor characteristics.

3.1 Physicochemical Properties

Inherent properties of chemicals drive potential for human and ecological exposure in the context of a given system. Mackay et al. (2001) argue that chemical substances can be classified with respect to exposure potential meaningfully and independently of quantity released to the environment according to their persistence, bioaccumulation, and potential for long-range atmospheric transport. Persistence, the propensity to remain in the environment for long periods of time, and bioaccumulation, accumulation in primary aquatic living organisms via the food

chain, were key factors used to conduct screening assessments of substances listed on the Canadian Domestic Substances List (DSL) (Meek, 2007). These factors are currently used by EPA to evaluate new chemicals before permitting their entry into the marketplace (Moss et al., 2000), and are important criteria for triggering requirements for exposure characterization under Europe's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation.

While persistence and bioaccumulation can be determined experimentally in the laboratory under highly controlled conditions, they are more often estimated computationally using a relatively small number of physicochemical properties (Zachary and Greenway, 2009). These estimates and the algorithms on which they are based, however, are not without limitations. Indeed, model selection and the assumptions chosen in their application may result in different rankings of human exposure potential (MacDonald et al., 2002; Arnot et al., 2010). Chemical characteristics of persistent organic pollutants are well established, namely, low water solubility, high lipid solubility, semi-volatility, and high molecular mass. Persistence and bioaccumulation are useful for assessing exposure via contact with the surrounding natural outdoor environment and chronic exposure through certain types of food (fish, for example), but they provide little information on exposure in the indoor environment. To date, the effort applied toward the identification of physical and chemical properties that most directly influence indoor persistence has been modest. Weschler and Nazaroff (2008) identified molecular weight, melting point, saturation vapor pressure, Henry's law constant, water solubility, octanol–water partition coefficient, and octanol–air partition coefficient as directly influencing the partitioning and transport of semi volatile organic compounds in indoor environments. However, many screening models rely only on vapor pressure and product characteristics (ECHA, 2010).

3.2 Product Characteristics (Manufacture, Formulation, Use, Lifecycle)

The general population is typically exposed to chemicals indirectly through the environment or directly through use and contact with consumer products. As such, data are required to characterize potential for human exposure to a given chemical as a function of relevant product characteristics and across the chemical/product lifecycle. Because the vast majority of chemicals are not monitored in environmental media, production volume is often the primary surrogate by which chemicals are evaluated for potential exposure. Though many legacy environmental contaminants were high production volume chemicals, the assumption that human exposure is strongly correlated with production volume has been questioned (Muir and Howard, 2006).

Direct and indirect exposure to chemicals in products typically exceeds environmental exposure, except perhaps for those residing in the vicinity of an emission site (Wormuth et al., 2007). Health Canada was legislatively mandated to categorize by exposure all chemicals in commerce listed on their DSL. The DSL categorization exercise suggested that persistence and bioaccumulation potential did not necessarily reflect human exposure and highlighted the importance of direct consumer exposure relative to indirect environmental exposure. Specifically, use category was identified as an important surrogate for exposure-based prioritization, with use profiling proving far more influential than production volume in determining exposure rankings (Meek et al., 2007). At a more fundamental level, exposures to chemicals in products are commonly thought to be a function of the vapor pressure of the chemical and the characteristics and use scenarios associated with the products. For example, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Targeted Risk Assessment (TRA) consumer exposure model approved for Tier 1 exposure estimation under REACH provides

scenarios with defaults for contact factors for 46 specific types of products (e.g., furniture spray polish) and articles (e.g., plastic toys). The user needs only to provide the vapor pressure of the chemical, the concentration of the chemical in the product, the amount of product used per event, and the number of events per day (ECHA, 2010). Several other consumer exposure models are discussed in Kephelopoulos et al. (2007). Little work has been done to compare results of scenario-based modeling to measured concentrations of chemicals in indoor environments. Since information on product composition and consumer product use is required as inputs to these models, the need for product formulation and product use information has been identified as critical for assessing human exposure to chemicals (Wormuth et al., 2007).

3.3 Emission Characteristics

A more direct indicator of potential for human exposure than production volume and general use classification is emission of the chemical into the system of interest (e.g., the general environment). Emission data are required at the relevant scale for exposure analysis. Either a compound is released far-field from the receptor individual/population and information on environmental fate and transport is required, or the compound is released near-field and information on the fate and transport in the immediate (usually indoor) environment is required. In either case, data on mass emitted and media into which the chemical is emitted and/or the media from which the chemical is available for contact are required (Schneider et al., 2011; MacLeod et al., 2011). A lack of information on emission factors during the production of chemicals, the downstream use of chemicals in various products, and the emissions from products will limit effective evaluation of human exposure (van Leeuwen et al., 1996; Fryer et al., 2006).

3.4 Routes and Pathways

Data on chemical levels in environmental media (e.g., air, dust, water, food) combined with information on chemical movement among media provide critical information for estimating potential human exposure. While data on chemical levels in biological media (e.g., blood, urine, nails) provide a direct indication of human exposure, data on levels in environmental media can serve to identify the most effective means for reducing or preventing risk by blocking or eliminating exposure pathways.

3.5 Receptor Characteristics

Because the ACToR system is chemical specific, the scope of the exposure data identified and described within this analysis does not include direct consideration of receptor characteristics. Receptor here refers to human beings, not target organs or tissues, and receptor characteristics include time-location-activity patterns and contact rates. Other related resources focus on exposure factor data including EPA's Consolidated Human Activity Database (McCurdy et al., 2000) and EPA Exposure Factors Handbook (EPA, 2011). However, data related to use of specific chemicals in products designed for a vulnerable receptor population such as children were explored.

For this analysis we collated publicly available data to evaluate the data landscape (i.e., coverage, or number of chemicals) across these critical aspects of exposure space. A better characterization of data across this exposure space will aid in identifying key gaps required to augment the often putative surrogates (e.g., production volume) currently being used to represent exposure. As additional data sources are further identified and evaluated, the data can be tied back to these key aspects of the exposure space.

4. EXPOSURE DATA RESOURCES

Several sources of exposure-related data are publically available via the internet. Schwela and Hakkinen (2004) review websites offering a wide range of human exposure assessment information from governmental agencies, environmental organizations, and professional societies. The focus is on training materials, guidelines, and tools, rather than on chemical concentrations. Patterson et al. (2002) largely align their review along the four components of the National Academy of Sciences risk assessment paradigm and cover exposure-related data in two pieces: hazard identification and exposure assessment. They describe websites providing exposure assessment guidance and models as well as data related to chemical release, physicochemical properties, and environmental fate and transport from a broad range of organizations. Polifka and Faustman (2002) also present an overview of resources organized along the lines of the risk assessment paradigm but targeted towards health care professionals concerned with developmental toxicants and their effects. The resources evaluated under exposure assessment are related solely to occupational exposures. Felsot (2002) presents a comprehensive listing of pesticide databases that, while focused on toxicology, also point the reader towards available data on pesticide residue levels in food and drinking water. Stone and Delistraty (2010) present an inclusive listing of sources of exposure information in the supplemental materials to their article discussing identification of high priority chemicals under the Washington State Children's Safe Product Act.

Among the most comprehensive of several data aggregators that provide access over the internet to exposure-related values is the Hazardous Substances Data Bank (HSDB) on the National Library of Medicine (NLM) Toxicology Data Network (TOXNET) (NLM, 2011). The HSDB focuses on the toxicology of potentially hazardous chemicals with comprehensive, peer-reviewed toxicology data for over 5,000 chemicals. It also provides information on human

exposure, industrial hygiene, chemical and physical properties, manufacturing and use, and environmental fate. All data are abstracted from a core set of reference books, government documents, technical reports and selected primary journal literature and are peer-reviewed by a committee of experts. Similarly, NLM's Haz-Map links jobs and hazardous tasks with occupational diseases and their symptoms via expert review of publications.

EPA's Data Finder (<http://www.epa.gov/datafinder/>) site indexes EPA data sources and is organized by topic: air, climate change, health risks, pollutants and contaminants, waste, and water. A chemical specific search returns a list of data sources, each with a basic overview including geographic scale and other contextual information. Where available, access is provided to the data source, offering access to data in downloadable formats. Another EPA resource is the Envirofacts Data Warehouse (<http://www.epa.gov/enviro/>). It is a single point of access to several EPA databases with environmental data for any geographical location in the United States (i.e., specific ZIP Code, City or County and State). The website provides information on industrial activities that may affect air, water, and land and allows the user to generate maps of environmental information, including toxic chemical releases, water discharge permit compliance, hazardous waste handling processes, Superfund status, and air emission estimates.

EPA's ACToR information management system contains publically available data on chemicals in the environment, including information on hazard, exposure, risk assessment, regulations and production. All exposure-related data in ACToR have been organized into categories based on the taxonomy presented in **Table 1**. The relationship among the taxonomy categories is displayed in a network diagram in **Figure 1**, showing the hierarchical relationship among the assay categories. All publically available exposure-related data for chemicals listed by CAS number that the ACToR team uncovers are brought into the system, but no proprietary data

are included. The types of exposure data can be classified into the following categories: (a) Production/import volumes and processes, (b) Consumer product ingredients and usage, (c) Environmental release, (d) Environmental media concentrations, and (e) Human exposure/biological monitoring. The following is a listing of some of the key data sources. The list is not meant to be comprehensive, and a complete listing of data sources can be found in the “Data Collections” section of ACToR.

Information on production and import volumes in ACToR includes all non-confidential information from the EPA Inventory Update Reporting (IUR) Program, which collects screening-level, exposure-related information on chemical substances produced at any single site at 25,000 lb or more. Information is also captured, albeit with a lag, from the EPA High Production Volume Information System. Efforts are underway to establish a direct link to the EU European Substances Information System, and additional production process information can be accessed through links provided to the ATSDR Toxicological Profiles.

Industrial use and product ingredient information comes from, among other sources, the U.S. Food and Drug Administration (FDA) Everything Added to Food in the United States (EAFUS) inventory, FDA’s Approved Drug Products, and Health Canada’s Substances in Cosmetics and Personal Care Products Regulated Under the Food and Drugs Act listing. It should be noted that the NLM Household Products Database is not included in ACToR due to licensing issues, and several publically accessible databases that link only through product names are also excluded; these include SC Johnson’s What’s Inside, Clorox’s Ingredients Inside, and the Environmental Working Group’s Skin Deep Cosmetic Safety Database. Efforts are currently underway to access Walmart’s online MSDS search engine for information on product ingredients and the Substances in Products in the Nordic Countries (SPIN) database

(<http://www.spin2000.net>) for more detailed information on industrial uses and product categories of chemicals.

Environmental release data comes from a number of sources including the EPA's Toxics Release Inventory and National Emissions Inventory, the EPA Pesticide Industry Sales and Usage Market Estimates, the Great Lakes Commission Toxic Air Emissions Inventory, and the Health Canada National Pollutant Release Inventory. Environmental fate information comes from the EPA Toxic Substance Control Act Test Submission Database, EPA Inert Ingredients in Pesticide Products tolerance reassessment decisions, the U.S. National Oceanic and Atmospheric Administration Pharmaceuticals in the Environment Information for Assessing Risk database, and the EPA ECOTOX database. There is a noticeable lack of available data on indoor emissions from both building materials and consumer products; such information would be particularly useful for assessing direct exposures to compounds formulated into consumer products and articles.

Sources of measured environmental concentrations include EPA's National Contaminant Occurrence Database for unregulated contaminants in public water systems, EPA AirData for ambient air monitoring data, EPA National Air Toxics Assessment for modeled ambient data, UN Intergovernmental Panel on Climate Change Greenhouse Gas Inventories, U.S. Department of Agriculture (USDA) Pesticide Data Program (PPD) database of pesticide residues in the U.S. food supply, summaries of the FDA Total Diet Study market basket survey of contaminants in food, EPA Human Exposure Data System (HEDS) for exposure-related media concentrations from observational studies, and occupational exposure limits from OSHA and NIOSH. Human Exposure Monitoring data includes personal air, dermal, and biomarker values from

observational studies contained in EPA HEDS, and biological monitoring from the CDC National Health and Nutrition Examination Survey (NHANES)

5. ANALYSES USING ACToR

5.1 Exposure and Use Taxonomy

In order to illustrate the types and amount of exposure data publicly available (**Table 1**), we organized analyses around three classes of chemicals using queries of the ACToR database. The first are industrial chemicals covered by the EPA IUR database, combining all chemicals that had been on the IUR list in any reporting period from 1986 to the present (14,609 total chemicals or unique CASRN). The second list includes chemicals that may be lawfully added to food, which is comprised of the substances on the FDA EAFUS list (2,820 chemicals or unique CASRN). The third list contains chemicals for which there is concern for exposure to children (714 unique chemicals or CASRN), and is derived from a number of studies and lists, including: California Air Resources Board Children's School Bus Exposure Study; EPA Voluntary Children's Chemical Evaluation Program (VCCEP); EPA Children's Total Exposure to Persistent Pesticides and Other Persistent Organic Pollutants study; German Environmental Survey; HUD First National Environmental Health Survey of Child Care Centers; and EPA Minnesota Children's Pesticide Exposure Study. For comparison, we also included a column ("Toxicity Data Landscape") showing the number of chemicals previously evaluated for availability of toxicity data. The rows of **Table 1** represent major categories in the assay taxonomy related to exposure and use categories. Each row lists the number of chemicals with data in that category in the ACToR, IUR, Food Use, and Children's Exposure chemical sets as well as those with corresponding toxicity data. The second row of **Table 1** presents the highest

level of the hazard taxonomy, and the third row presents the overall summary of publicly available chemical-specific exposure-related data, the highest level of the Exposure taxonomy (**Figure 1**). There are a total of 21,527 chemicals with some form of exposure-related data. Of these, 14,566 are industrial chemicals reported under the IUR at least once since 1986; 2,485 may be added to food in the U.S.; and 203 have been identified as chemicals of concern to children.

5.2 Chemical Coverage Based on Properties

As an example of an analysis of key aspects of exposure space, we have examined two physicochemical properties of particular interest, bioaccumulation (quantified by the BCF) and biopersistence (quantified by the biodegradation potential) for chemicals allowed in food and chemicals previously identified as being of concern for children's exposure. High values of log BCF indicate the potential to bioaccumulate in tissue, and low values of the biodegradation potential indicate the potential to be biopersistent. Examining all 5,489 chemicals that have been on the IUR since 1986 and for which EPISuite™ was able to calculate these two parameters, we find mainly polychlorinated biphenyls (PCBs) and organo-tin compounds of most concern for bioaccumulation and persistence, with log BCF > 4 and biodegradation potential < 0. Among the chemicals of concern for children's exposure in this region are 18 PCBs and two polycyclic aromatic compounds. There are no EAFUS compounds meeting both criteria, but several have a biodegradation potential < 0, mainly cyclic and linear poly-ethers, some of which are surfactants or solvents.

5.3 Historical Survey of Production Volumes

The EPA has been tracking production volume through the TSCA IUR since 1986. Every 4 years, manufacturers have been required to report import and manufacturing volumes of chemicals above thresholds ($>10,000 \text{ lb y}^{-1}$ per site until 2006, when the threshold was raised to 25,000 lb). **Figure 2** shows the number of chemicals in each volume category in each of the reporting years, showing that the categories are relatively stable over time. An exception is the significant drop in the number of chemicals under 500K lb from 2002 to 2006; this is principally due to the increased reporting threshold, so that the lowest volume chemicals were no longer reported. Despite an approximately ten-fold increase in global chemical production volume between 1970 and 2010 (IPCS, 2010), the number of medium and high production chemicals has stayed remarkably consistent. This suggests that the increase in chemical production is not necessarily due to manufacture of a greater number of chemicals each year, but instead due to increasing quantities of a relatively constant number of industrial chemicals. Several limitations of the IUR data have been identified (Denison, 2009), including that the thresholds for reporting apply to individual facilities, that certain categories of chemicals are excluded, and that the limited timeframe may not give an accurate picture of general production rates since the information applies only to the one year preceding the reporting year. Several limitations have been addressed in new reporting requirements implemented for the next cycle.

5.4 Overlap in Chemical Coverage: from Production Volume to Biomarker

The number of unique chemicals in ACToR with information on production volume, use category, chemical release, and measurement of concentrations in food, water, soil, air, and biological fluids is presented in **Figure 3**. Production volume information is available for 14,591 chemicals and use categories for approximately 2400 chemicals. Only a much smaller fraction has readily accessible measured concentrations in exposure-related media. Water is the medium

with the largest number of measured chemicals with over 1150 unique compounds. Measurement of chemicals in water is performed routinely by two separate agencies (EPA and U.S. Geological Survey), perhaps leading to the relatively large number. Nearly 800 substances have available measurements in soil, 720 are measured in food, and 670 are measured in air. Biomarker measurements are available for nearly 400 compounds, with over 200 measured in NHANES alone.

Figure 4 presents the overlap among chemicals with production volume information, use category, and various exposure media and biomonitoring concentrations (same categories as in **Figure 3**). Of the roughly 21,500 unique chemicals with exposure-related data in ACToR, only about 2500 are simultaneously found in more than one category. Less than 700 are found in three categories, and about 200-300 are found in either four, five, or six categories. Clearly the coverage across different types of exposure-related information is poor.

5.5 Chemicals of Concern to Children

Of the chemicals included in ACToR, approximately 700 have been identified as being of concern to children (**Table 1**). Exposure-related information exists for only 185 of these, with biomonitoring data available for even fewer. This small number of chemicals in this category highlights one important gap in the exposure data landscape – namely a well-documented listing of chemicals to which children are routinely exposed, along with the expected exposure scenarios or routes.

6. DISCUSSION

Expanding the scientific understanding of human exposure requires that existing human exposure data become more easily accessible. This analysis of readily available exposure data, spanning across the source-to-dose continuum from production volume to biological markers, reveals that for other than production volume and cursory use categories, the number of unique chemicals for which we currently have measured exposure-related media concentrations is low. This stands in contrast to the number of chemicals with available toxicity information, as a recent investigation of environmental chemicals being considered for the EPA ToxCast™ screening and prioritization program found 6,551 of 9,912 to have at least limited toxicity summaries available (Judson et al., 2009).

Our analysis addresses numbers of chemicals but does not consider the scale of the monitoring efforts. For example, although the number of chemicals for which we have ambient air measurements is low, criteria air pollutants (as defined by the Clean Air Act) are measured for compliance with the National Ambient Air Quality Standards (often on an hourly basis) by a monitoring network of thousands of stations that spans the entire country. Similarly, over 90 contaminants are regularly measured in some 26,000 public drinking water systems throughout the country. The more than 200 environmental chemicals measured in blood, serum, or urine as part of NHANES typically have a sample size of about 2000, are nationally representative, and are repeated every two years. On the other hand, many contaminants of indoor environments measured in proximal exposure media (e.g., house dust, indoor air) in observational exposure studies with publically accessible data have been measured only a handful of times with little or no attempt at achieving a representative sample of the larger population. The proximal exposure measurements in ACToR do not represent all of the exposure measurements available in the

scientific literature since at the present time there is no source of curated human exposure measurements with broad coverage in chemical and media space.

Consolidating exposure data into ACToR makes access to chemical-indexed data from disparate databases straightforward and allows for improved linkages with toxicity data. The rise of informatics, and the creation of increasingly advanced bioinformatics tools, is driving ever more sophisticated data analysis. The type of consolidation described herein facilitates an application of environmental informatics tools to exposure data. Application of large-scale, multidimensional analysis of the collocated toxicology and exposure data holds the promise of uncovering heretofore unrecognized environmental determinants of disease. It would also allow for establishing linkages with product usage data. The establishment of standards for the reporting and representation of exposure-related data would likely accelerate and broaden these efforts.

Advancements in standards for reporting and the representation of exposure-related information would also serve to facilitate data sharing, thereby widening the breadth of available exposure data. Exposure data standards would also reduce the ambiguity that impedes the understanding of human exposure by the broader science community. At a time when hand-interpretation of data is giving way to the era of automated interrogation of databases, standardization will facilitate linkages with other fields. Efforts, such as ACToR, to catalog and link the varied sources of exposure data hold great promise in supporting much needed predictive models for screening chemicals based on exposure.

In evaluating the data landscape, we were able to identify some of the strengths and limitations of the exposure data available for chemical risk assessment. The results suggest that

currently available exposure data are insufficient to provide the evidence base required to inform risk assessment and public health decision making. In this analysis of the 547,088 substances catalogued in the ACToR database, exposure information is readily available for only a small fraction. Even with new publically accessible data systems constantly being added with each new ACToR release, the total number of unique chemicals with exposure information is not expected to increase substantially, as we believe we are close to exhausting the world of large exposure-related datasets. Nonetheless, we are optimistic that efforts such as this will be useful in the development of exposure-based approaches for prioritizing chemicals for risk assessment. Furthermore, if enacted, proposed reform of U.S. legislation regulating potentially toxic chemicals (EPA, 2009a) would likely require chemical manufacturers to provide a minimum set of data, including more detailed data on exposure potential, for large numbers of chemicals. Such data requirements hold the promise of greatly expanding the universe of chemicals for which useful exposure-related information exists.

DISCLAIMER

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency review and approved for publication. The co-authors of this manuscript do not have any financial conflict of interest with any parties that were involved in this study.

ACKNOWLEDGMENTS

We thank Dr. Daniel Chang of the U.S. EPA for his very helpful review of this manuscript.

REFERENCES:

Applegate J, Baer K. 2006. Strategies for closing the data gap.

http://ecb.jrc.it/documents/Existing-Chemicals/PUBLIC_AVAILABILITY_OF_DATA/.

Arnot JA, Mackay D, Parkerton TF, Zaleski RT, Warren CS. 2010. Multimedia modeling of human exposure to chemical substances: the roles of food web biomagnifications and biotransformation. *Environ Toxicol Chem*. 29(1):45-55.

Brazma A, Krestyaninova M, Sarkans U. 2006. Standards for systems biology. *Nat Rev Genet* 7(8):593-605.

Chuprina A, Lukin O, Demoiseaux R, Buzko A, Shivanyuk A. 2010. Drug- and Lead-likeness, Target Class, and Molecular Diversity Analysis of 7.9 Million Commercially Available Organic Compounds Provided by 29 Suppliers. *J Chem Inf Model* 50(4):470-479.

Denison RA. 2009. Across the Pond: Assessing REACH's First Big Impact on U.S. Companies and Chemicals. Environmental Defense Fund, New York.

Dobson CM. 2004. Chemical space and biology. *Nature*. 432(7019):824-8.

ECHA (European Chemicals Agency). 2010. Guidance on information requirements and chemical safety assessment Chapter R.15: Consumer exposure estimation. ECHA-10-G-03-EN. Helsinki, Finland.

EPA. 2011. Exposure Factors Handbook 2011 Edition (Final). U.S. Environmental Protection Agency, Washington, DC. EPA/600/R-09/052F.

EPA. 2009. Essential Principles for Reform of Chemicals Management Legislation. Available: <http://www.epa.gov/oppt/existingchemicals/pubs/principles.html>

EPA. 2008. 2006 Inventory Update Reporting: Data Summary. Office of Pollution Prevention and Toxics. EPA/740/S-08/001.

- EPA. 1998. Chemical hazard data availability study. Office of Pollution Prevention and Toxics.
- Felsot AS. 2002. WEB resources for pesticide toxicology, environmental chemistry, and policy: a utilitarian perspective. *Toxicology*. 173(1-2):153-66.
- Fryer M, Collins CD, Ferrier H, Colvile RN, Nieuwenhuijsen MJ, Human exposure modelling for chemical risk assessment: a review of current approaches and research and policy implications. *Environ Sci Policy*. 9(3):261-274.
- Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect*. 117(5):685-95.
- Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, Cathey T, Transue TR, Spencer R, Wolf M. 2008. ACToR--Aggregated Computational Toxicology Resource. *Toxicol Appl Pharmacol*. 233(1):7-13
- MacDonald D, Breton R, Sutcliffe R, Walker J. 2002. Uses and limitations of quantitative structure-activity relationships (QSARs) to categorize substances on the Canadian Domestic Substance List as persistent and/or bioaccumulative, and inherently toxic to non-human organisms. *SAR QSAR Environ Res*. 13(1):43-55.
- Mackay D, McCarty LS, MacLeod M. 2001. On the validity of classifying chemicals for persistence, bioaccumulation, toxicity, and potential for long-range transport. *Environ Toxicol Chem*. 20(7):1491-8.
- MacLeod M, von Waldow H, Tay P, Armitage JM, Wöhrnschimmel H, Riley WJ, McKone TE, Hungerbühler K. 2011. BETR global - A geographically-explicit global-scale multimedia contaminant fate model. *Environ Pollut*. 159(5):1442-5.

- McCurdy T, Glen G, Smith L, Lakkadi Y. 2000. The National Exposure Research Laboratory Consolidated Human Activity Database. *J Exp Anal Environ Epi* 10:566-578.
- Meek ME, Armstrong VC. 2007. The assessment and management of industrial chemicals in Canada. In: Van Leeuwen CJ, Vermeire TG (eds.). *Risk Assessment of Chemicals: An Introduction*, 2nd Ed. Springer, Dordrecht, The Netherlands.
- Moss KT, S. Boethling RS, Nabholz JV, Auer CM. 2000. U.S. Environmental Protection Agency New Chemicals Program PBT Chemical Category. In: Lipnik RL et al. (Eds). *Persistent, Bioaccumulative, and Toxic Chemicals II*. American Chemical Society Symposium Series 773:138-150.
- Muir DCG, Howard PH. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. *Environ Sci Technol*. 40:7157-7166.
- Murmann JP. 2003. Chemical Industries after 1850. In: Mokyr J (Ed.). *The Oxford Encyclopedia of Economic History*. New York, Oxford University Press.
- NLM. 2011. Fact Sheet: Hazardous Substances Data Bank (HSDB[®]). Available: <http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html>
- Ogden T. 2009. Data sharing, Federal Rule of Evidence 702, and the lions in the undergrowth. *Ann Occup Hyg*. 53(7):651-5.
- Patterson J, Hakkinen PJ, Wullenweber AE. 2002. Human health risk assessment: selected Internet and world wide web resources. *Toxicology*. 173(1-2):123-43.
- Polifka JE, Faustman EM. 2002. Developmental toxicity: web resources for evaluating risk in humans. *Toxicology*. 173(1-2):35-65.

- Schneider T, Brouwer DH, Koponen IK, Jensen KA, Fransman W, Van Duuren-Stuurman B, Van Tongeren M, Tielemans E. 2011. Conceptual model for assessment of inhalation exposure to manufactured nanoparticles. *J Expo Sci Environ Epidemiol*. [Epub ahead of print]
- Schwela D, Hakkinen PJ. 2004. Human exposure assessment resources on the World Wide Web. *Toxicology*. 198(1-3):169-76.
- Stone A, Delistraty D. 2010. Sources of toxicity and exposure information for identifying chemicals of high concern to children. *Environ Impact Assess Rev*. 30(6):380-387.
- Van Leeuwen CJ, Bro-Rasmussen F, Feijtel TCJ et al. Risk assessment and management of new and existing chemicals, *Environ Toxicol Pharmacol*. 2(4):243-299.
- Weschler CJ, Nazaroff WW. 2010. SVOC partitioning between the gas phase and settled dust indoors. *Atmos Env*. 44(30):3609-3620,
- Weschler CJ. 2009. Changes in indoor pollutants since the 1950s. *Atmos Env*. 43:156–172.
- Wilson MP, Schwarzman MR. 2009. Toward a new U.S. chemicals policy: rebuilding the foundation to advance new science, green chemistry, and environmental health. *Environ Health Perspect*. 117(8):1202-9.
- Wormuth M, Demou E, Scheringer M, Hungerbühler K. 2007. Assessments of direct human exposure: the approach of EU risk assessments compared to scenario-based risk assessment. *Risk Anal*. 27(4):979-90.
- Zachary M, Greenway GM. 2009. Comparative PBT screening using (Q)SAR tools within REACH legislation. *SAR QSAR Environ Res* 20(1-2):145-57.

Table 1: Mapping of chemicals to exposure taxonomy categories. Number of chemicals with exposure-related information and in categories of particular interest.

CATEGORY 1	ACToR	IUR	FOOD USE	CHILDREN'S EXPOSURE	TOXICITY DATA LANDSCAPE
Total	547088	14609	2820	714	17347
Hazard	108443	10891	2605	196	13062
Exposure	21527	14566	2485	203	16138
Exposure; Biomonitoring	390	106	177	143	242
Exposure; Chemical Release	1737	898	802	140	1599
Exposure; Chemical Release; Accidental	155	133	52	10	154
Exposure; Chemical Release; Media; Air	225	180	112	59	225
Exposure; Environmental Fate	7023	3333	1179	163	4101
Exposure; Group; Age	1870	695	462	194	1057
Exposure; Group; Age; Adult	51	6	50	19	44
Exposure; Group; Age; Child	1868	695	460	194	1055
Exposure; Group; Ecological	872	155	484	65	597
Exposure; Group; Vulnerable Population	256	181	79	54	207
Exposure; Media	3716	1385	1923	201	2464
Exposure; Media; Air	670	434	349	161	622
Exposure; Media; Air; Ambient	43	42	20	13	43
Exposure; Media; Air; Residential	69	65	29	40	65
Exposure; Media; Air; Residential; Indoor	67	64	28	38	64
Exposure; Media; Air; Residential; Outdoor	28	27	16	28	27
Exposure; Media; Air; Residential; Personal	33	31	19	33	31
Exposure; Media; Air; Residential; Personal; In Transit	17	16	11	17	16
Exposure; Media; Biosolids	43	18	21	10	21
Exposure; Media; Sediment	166	76	131	106	139
Exposure; Media; Soil	788	396	390	160	673
Exposure; Media; Surfaces	36	4	34	36	34
Exposure; Media; Water	1153	484	473	134	783
Exposure; Media; Water; Drinking	299	122	125	63	179
Exposure; Media; Water; Ground	608	316	315	103	532
Exposure; Media; Water; Surface	188	85	130	59	152
Exposure; Media; Water; Waste	90	44	51	18	56
Exposure; Production	15635	14553	1330	150	14954
Exposure; Production; Disposal	813	465	357	102	740
Exposure; Production; Export	832	352	81	26	383

CATEGORY 1	ACToR	IUR	FOOD USE	CHILDREN'S EXPOSURE	TOXICITY DATA LANDSCAPE
Exposure; Production; Import	36	32	7	2	32
Exposure; Production; Site	6200	6200	833	68	6200
Exposure; Production; Transport	677	437	325	94	663
Exposure; Production; Volume	14591	14551	1133	90	14554
Exposure; Source	6927	6403	1237	137	6848
Exposure; Source; Agricultural	541	177	404	58	492
Exposure; Source; Industrial	639	429	313	81	630
Exposure; Source; Residential	27	2	24	27	23
Exposure; Source; Workplace	6297	6266	865	69	6279
Use Category	7587	2416	2093	174	3810
Use Category; Consumer	3483	955	565	55	1156
Use Category; Consumer; Children	1732	698	362	85	983
Use Category; Consumer; Cosmetics	185	73	23	0	85
Use Category; Consumer; Fabrics, Textiles	117	112	37	6	114
Use Category; Consumer; Pharmaceutical	2446	324	307	21	379
Use Category; Consumer; Soaps and detergents	173	154	68	1	158
Use Category; Food	2466	904	1713	146	1658
Use Category; Food Contact	21	11	21	0	13
Use Category; Occupational	1234	988	469	77	1107
Use Category; Pesticides	3320	1200	1748	150	2378

LIST OF FIGURES

Figure 1: Network of exposure taxonomy used in ACToR. Certain terms are reused in different branches of the taxonomy leading to the use of incremental indices (.1, .2, etc.)

Figure 2: Production over time of MPV and HPV chemicals based on 1986-2006 IUR by volume category.

Figure 3: Number of unique chemicals by data type in ACToR.

Figure 4: Number of chemicals for which one or more taxonomy categories of exposure-related information are available in ACToR. A chemical can be tied to a maximum of eight categories: five categories with measured concentrations (air, soil, food, water, biomonitoring) and three categories of exposure surrogates (production volume, use category and chemical release).

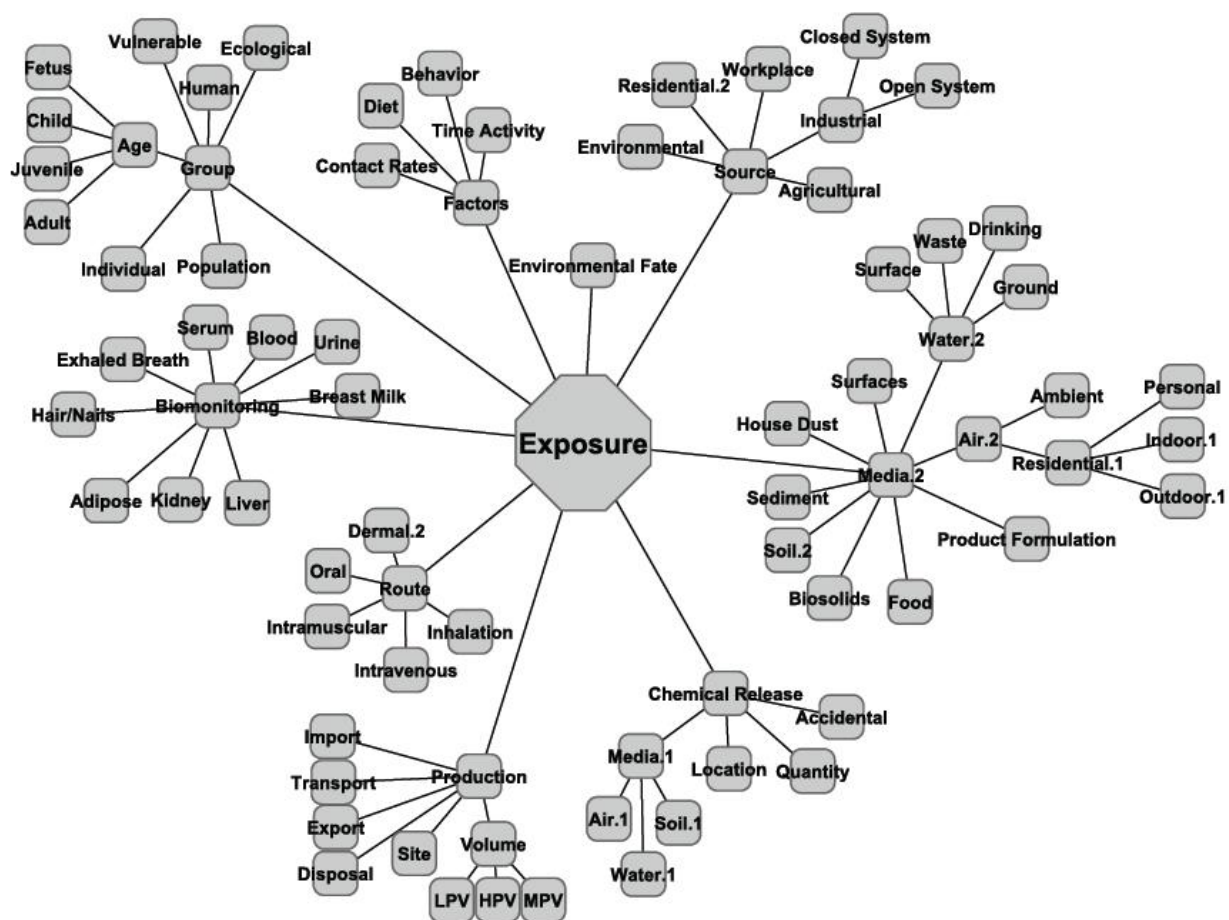


Figure 1: Network of exposure taxonomy used in ACToR. Certain terms are reused in different branches of the taxonomy leading to the use of incremental indices (.1, .2, etc.)

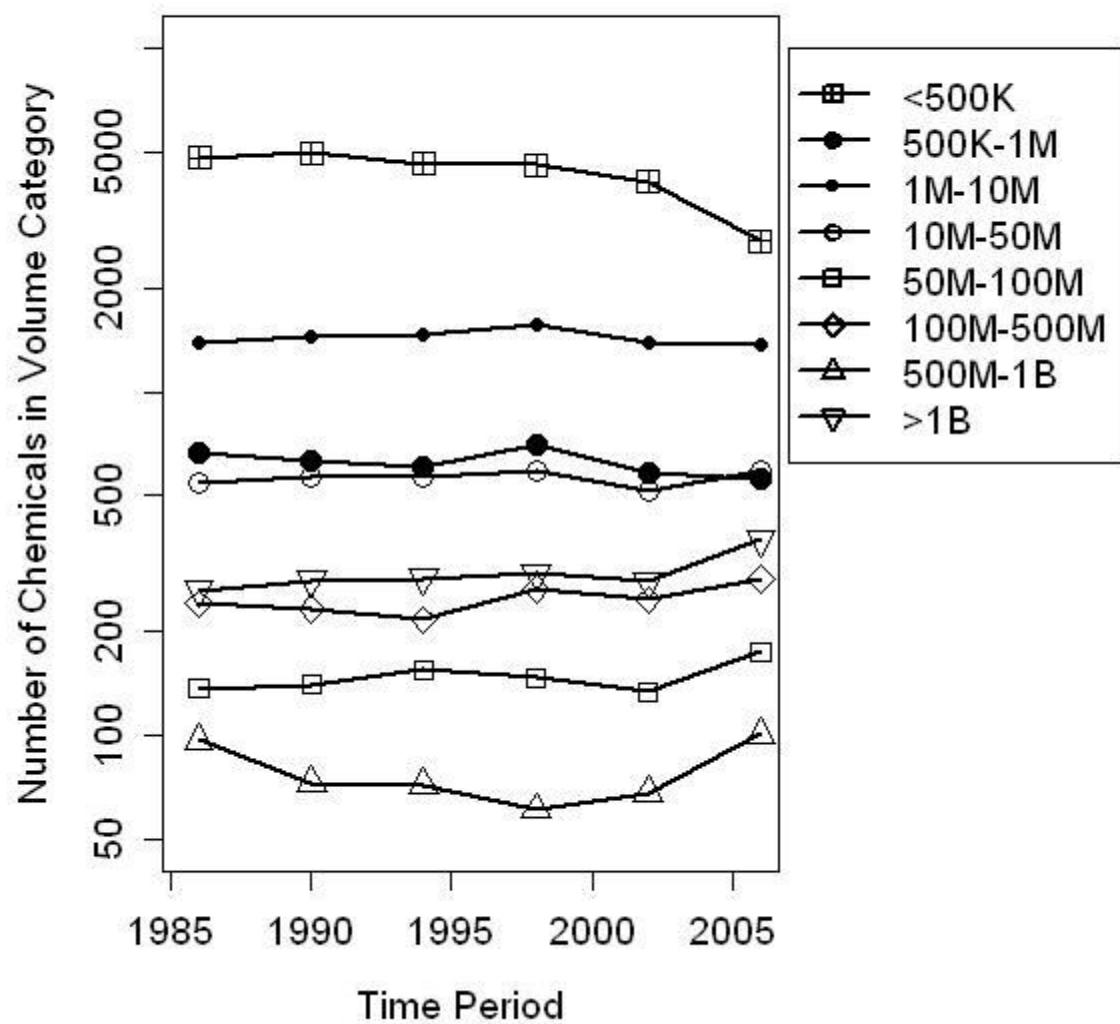


Figure 2: Production over time of MPV and HPV chemicals based on 1986-2006 IUR by volume category.

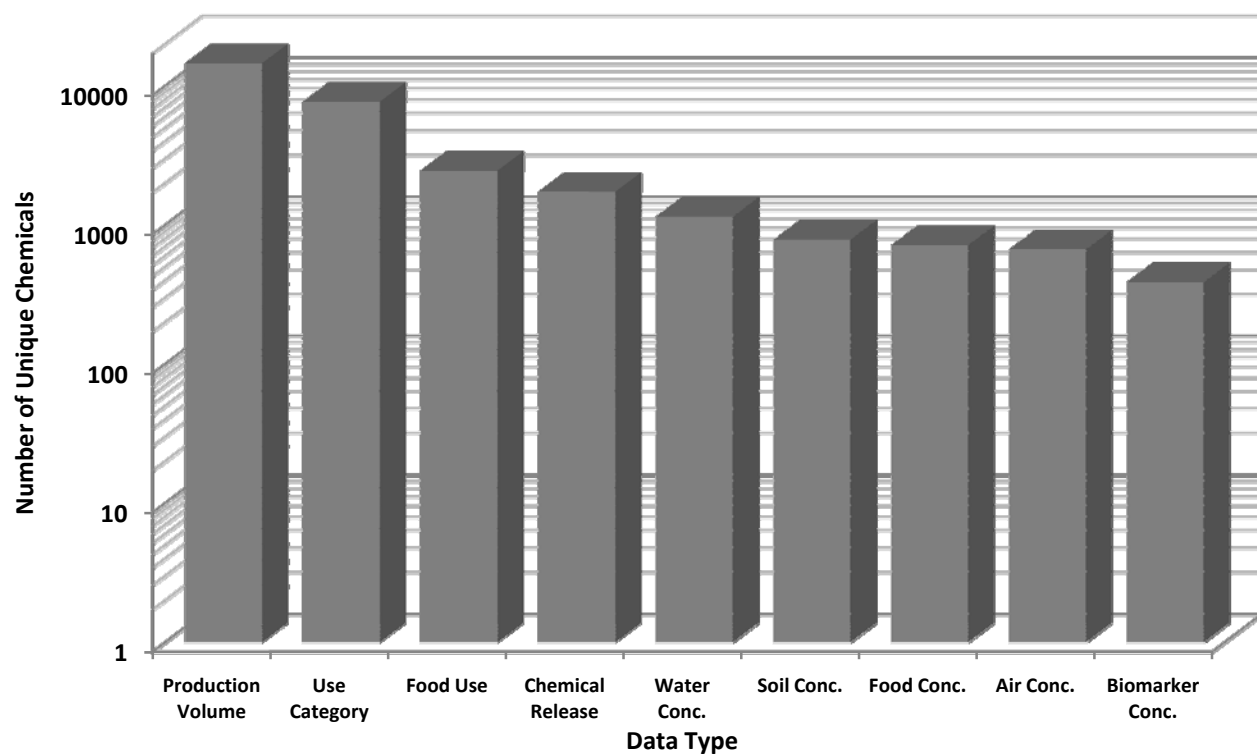


Figure 3: Number of unique chemicals by data type in ACToR.

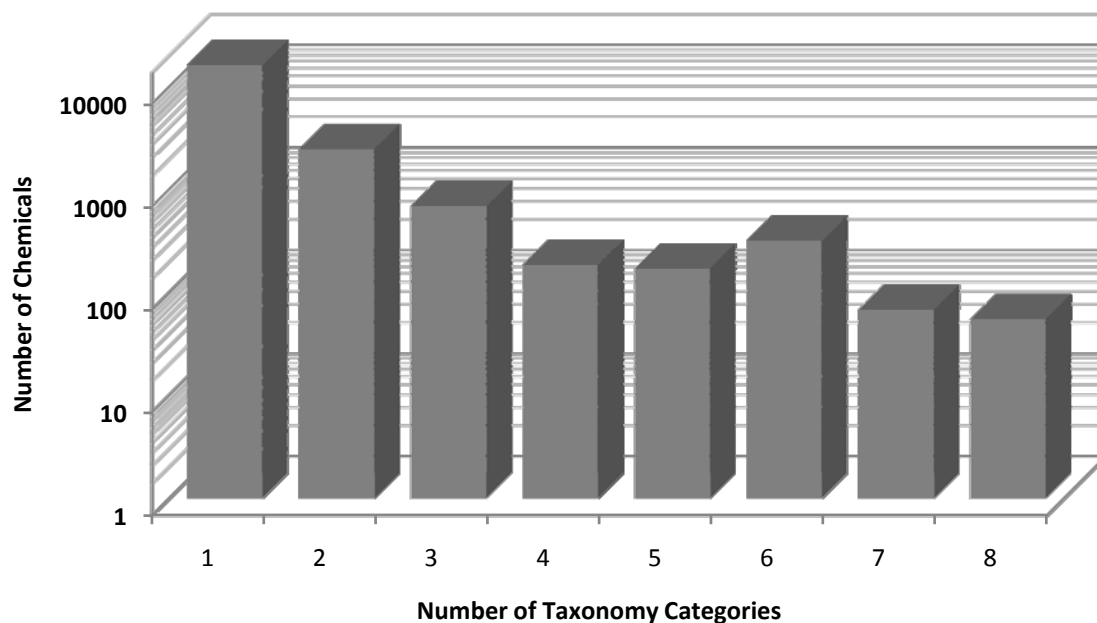


Figure 4: Number of chemicals for which one or more taxonomy categories of exposure-related information are available in ACToR. A chemical can be tied to a maximum of eight categories: five categories with measured concentrations (air, soil, food, water, biomonitoring) and three categories of exposure surrogates (production volume, use category and chemical release).