
CCL Universe Example Data Set- Update

Report for the NDWAC CCL Work Group
Plenary Meeting
September 17-18, 2003

Updates to the CCL Universe Example Data Set

- At the July 16-17 meeting, NDWAC requested additional data sources be added to the Data Set to see if affects Gates Analysis
- This update reports on activities since July 16-17
 - Added two occurrence data sources to supplement Universe Data Set – 25 currently
 - Updated classification of data element types
 - Lists called Occurrence Information
 - Redefined many data elements from “other” to data or information
 - Updated Gates analyses and summary statistics

Additional Data Sources

- Addition of occurrence data sources
 - United States Geological Survey (USGS) - National Water Quality Assessment (NAWQA) Program
 - List of chemicals as placeholder
 - Will update with data from USGS
 - USGS - National Reconnaissance of Emerging Contaminants (NREC)
 - Summary Statistics
 - Hazardous Substances Data Bank (HSDB)
 - – in progress

Classifying Data Elements

- Reclassification of data element types

- Data elements are classified based on previous NDWAC Gates Classification:
 - Health Effects data
 - Health Effects information
 - Occurrence data
 - Occurrence information
 - “Other”
- Replaced some “other” classifications

Examples of Reclassifications

Data Source	Data Element	Prior Classification	New Classification
HPVMS_Master Summary Table	ACUTE	Other	HE INFO
HPVMS_Master Summary Table	CHRONIC	Other	HE INFO
HPVMS_Master Summary Table	TRI_HIGH	Other	OCC INFO
HPVMS_Master Summary Table	TRI	Other	OCC INFO
ITER_ATSDR Noncancer Oral	Basis(EXP) (mg/kg-day)	Other	HE DATA
ITER_ATSDR Noncancer Oral	Basis(ADJ) (mg/kg-day)	Other	HE DATA

Data Set Findings: Overview

- Updated Findings of the Example CCL Universe Data Set
 - Summary data set statistics
 - Number of chemicals with health effects and occurrence data and information
 - Number of chemicals with only health effects or occurrence data *or* health effects or occurrence information
- Evaluating other sources requiring text file extraction

Findings: Data Set Statistics

■ Data sources

- Example data set contains data from 25 data sources
- Data is contained in 89 tables
- Approximately 30,500 records
- 11,128 unique chemicals

■ Data elements

- Tables contain 862 data elements (fields)
- Data set contains 150 health effects and occurrence data elements
- 211 chemicals have health effects and occurrence data

Example CCL Universe Data Set Summary

Statistics (compared with previous set)

Database Subset	Data Sources	Number of Chemicals	Data Elements - Available Data or Info Only	Data Elements - Total (data, info, other)
Entire Universe Data Set	25 (23)	11,128 (10,360)	372 (293)	862 (848)
All Chemicals with demonstrated or potential Health Effects AND demonstrated or potential Occurrence	20 (18)	1,741 (774)	372 (293)	862 (848)
Chemicals with demonstrated Health Effects AND demonstrated Occurrence	18 (16)	211 (62)	150 (157)	<862 (<848)

Number of Chemicals with Data/Information

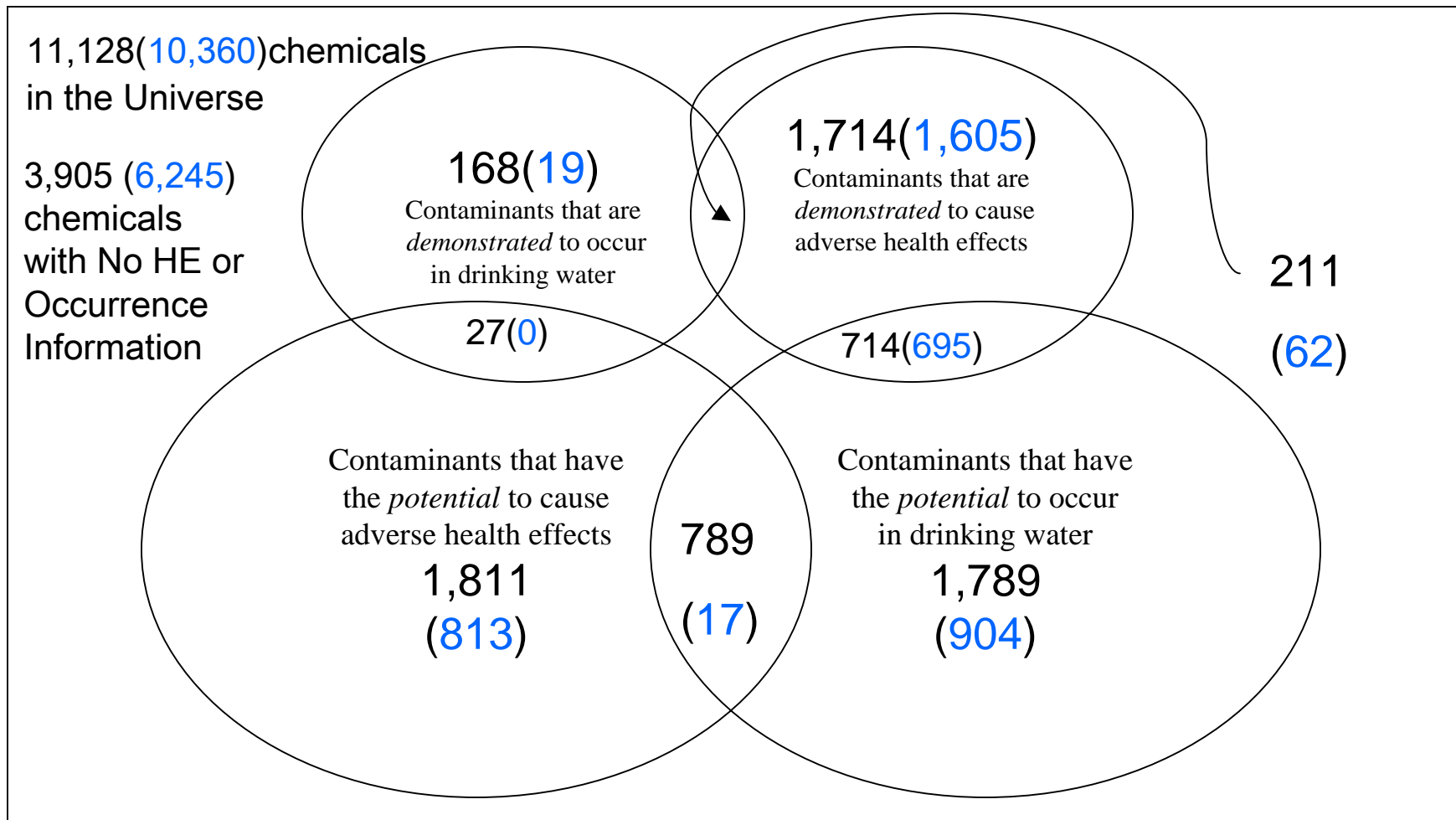
- Summary update on the number of unique chemicals in the Example CCL Universe Data Set
- The counts are related to the reclassification of data elements from “Other” to “Data” and “Information,” previously discussed by the NDWAC and EPA work groups

Number of Chemicals with Data/Information

(Compared with previous data set from 23 sources)

Category or Criteria of Data/Information Available; Gate Screening Criteria	Number of Chemicals in Example Data Set Meeting the Criteria	Percentage of Chemicals in Example Data Set Meeting the Criteria
Health Effects DATA AND Occurrence DATA; Gate 1	211 (62)	1.89% (0.6%)
Health Effects Information AND Occurrence DATA; Gate 2	27 (0)	0.24% (0.0%)
Health Effects DATA AND Occurrence Information; Gate 3	714 (695)	6.42% (6.7%)
Health Effects Information AND Occurrence Information; Gate 4	789 (17)	7.09% (0.2%)
Subtotal of Candidates Through Gates 1-4:	1,741 (774)	15.6% (7.5%)

The CCL Universe Example Data Set and the NRC'S Venn Diagram



Lessons Learned/Next Steps on CCL Universe Data Sources

- Reclassification/additional occurrence data doubled the number of chemicals meeting the “Gate” criteria
 - This will not continue indefinitely
 - Broader definitions of screening information affects gates
- Currently the process appears to be limited by occurrence data
- Next steps include:
 - Evaluating additional sources
 - Evaluating the quality of sources
 - Test screening approaches